Jun 3rd, 9:00 AM - Jun 6th, 5:00 PM

Probative Inference

Michael Scriven
Claremont Graduate University

Probative Inference

MICHAEL SCRIVEN[1]

Claremont Graduate University
150 E. 10th Street
Claremont
CA 91711
USA
scriven@aol.com

Terminology: I take the term 'probative' from jurisprudence, where it means 'having the quality or function of proving or demonstrating' a conclusion,[2] presumably by contrast with 'supporting, suggesting, or indicating' a conclusion, and I use the term with that meaning, except that I use it by contrast with two other established processes for doing the same task, notably deduction, and induction in a special sense of the latter. The special sense is intended to preclude using 'inductive' to cover all non-deductive reasoning, and refers to the type of induction for which we have some kind of logical analysis, namely statistical inference from samples to population[3]. I make no claim about the relation of 'probative' to several other terms that have been coined to refer to legitimate inference types other than deduction, which usually include some kinds of induction. These include logics called defeasible, non-monotonic[4], default, autoepistemic[5], paraconsistent and relevance (hope springs eternal!): reasoning called plausible, prima facie, presumptive, analogical (as John Wisdon suggested), conductive (Wellman's term, this and the previous one being just two of Blair's many suggestions for inclusion here), and of course Toulmin's rebuttable, Perelman's argumentative, and Peirce's abductive reasoning; since I find it hard to get a clear and generally acceptable definition of these terms, or, with the neologisms, to find them getting any substantial traction on the ground of real arguments. But I do think that they were all coined from the same sense of a need to identify a cate-

---

[1] For acknowledgments, see endnote.

[2] *New Oxford American Dictionary*. *Merriam-Webster* has this as sense 2; sense 1 is "exploratory," which I think is a rare use and would regard as a connotation of probative in my sense.

[3] The inference from the cards in poker player X's hand when there is still one card to be dealt, to the conclusion that it has a 93% chance of being the best hand dealt is inductive in my sense; if we assume the pack is well-shuffled and fairly dealt, it's deductive; the inference from this and an expert player's prediction, knowing all the players well, that s/he'll win the hand is probative in my sense.

[4] In his major contribution to this literature, John Pollock takes 'defeasible' to mean 'non-monotonic' i.e., such that the addition of a premise can lead to the failure of the argument to establish the conclusion, without contradicting any other premise ("Defeasible Reasoning" in *Reasoning*, (J. Adler and L. Rips, eds., Cambridge, 2008), p. 453. But he includes perception, which I categorically distinguish from inference; and does not consider inferences from analogies, rough estimates, and approximations, i.e., those by most experts in practice, and by most scientists using most of the 'laws' in physics, astronomy, and the forensic sciences, which I regard as important types of probative inference.

[5] Default and autoepistemic logics are referenced by Hank Kyburg in *Reasoning*, op. cit., p. 293.

Scriven, M. (2009). Probative Inference. In: J. Ritola (Ed.), *Argument Cultures: Proceedings of OSSA 09*, CD-ROM (pp. 1-10), Windsor, ON: OSSA.

gory of legitimate inference other than the classic duo. Like many of them, I think there is still room to do more in the way of providing an informal structure in terms of which the homeless family of concepts like prima facie, ceteris paribus, presumptive, enthymemes, burden of proof, ostensive definitions, etc., can find spatial shelter.

Within the category of probative inference I distinguish two sub-types: logical (i.e., based on meaning rules) and empirical (based on factual (meaning non-evaluative) evidence and empirical generalizations). An example of the first would be the taxonomic (sortal?) inference from the presence of a number of the most important criteria for the identification of, say, an apple, and the conclusion that it is an apple. I refer to these cases as criterial inference. An example of the second type would be the inference from the data available in 1990 to the conclusion that heavy smoking often causes lung cancer. There are plenty of intermediate cases where for example: (i) some of the facts appealed to are not criteria but only indicators (i.e., empirically but not logically correlated with applehood); or (ii) the general premise is arguably but not indisputably meaning based, e.g., cases using the laws of motion, and laws about ideal types such as ideal gases or ideally elastic bodies or ideal markets. And there are plenty of cases where the inference involves both types of (stated or implicit) premise.

Given that many of the nouns for things in common or widespread scientific use are now cluster concepts—that is, defined in terms of an indeterminate number of differentially weighted criteria[6] rather than a synonymous phrase—however neatly defined they may have been when introduced; and given the preponderance of cases of empirical inference which are neither deductions from true generalizations nor statistical inferences from samples, it seems plausible to conclude that much, perhaps even most, of our commonplace inferences, in everyday life and in science, are probative inferences.[7] It therefore seems plausible to argue, as my first suggestion, that our texts should spend more time on the explication of probative inference. I here try to add a few paragraphs aimed in that direction.

The 'criterion→cluster concept' inference provides a weak kind of illustration of the main type of *logical* probative inference. What can we say about its empirical cousin? We might begin by asking if there is anything in common to the logical form of each. It appears that both are engaged in a species of pattern-recognition. The first is seeking a pattern within the repertoire of existing *concepts*: the second seeks a pattern within the repertoire of existing *explanation-types*. Now, I realize that talking about patterns is pretty sloppy stuff compared with the simple charms of the neo-positivist deductive model of explanation, even though that had to be loosened up into the inductive model with its embarrassing 'requirement of total evidence.' My only excuse is that I cannot think of anything more exact that is correct. Let me expand a little on what I think we can say, and see if that's enough sugar-coating to get you to swallow the pill. I'm going to do some talking about brain functions in a minute, and let me preface that by saying that I

---

6 Criteria are here taken to be the *general properties* that are typically provided in explaining the meaning of a term. These include necessary conditions, although they are comparatively rare with common concepts, in science or life; and exclude particular descriptions (literal ostensive definitions) and of course real ostensive definitions.

7 I have been chasing the clarification of criteria for some time (see for example, "The logic of criteria" in *Journal of Philosophy* (1959, pp. 857-868), and that of scientific probative inference even longer (*Explanations…* (D. Phil. thesis, Oxford, 1956, (UMI Microform) for the concept of 'selection-explanations,' a possible precursor of 'inference to the best explanation.')

find almost everything that is advertised as an insight into human behaviour obtained from neurology to be a charade of confusing correlation and causation. But I'm going to propose what I think is one of the exceptions. Cognitive neuroscience is not just a new jargon, though what I have to offer did not come from research done there.[8]

First, the heuristic here—it would be a gross exaggeration to call it an algorithm—is what I'm going to call the General Elimination Method, mainly because it has a memorizable acronym.[9] The brain, whether it is managing a tacit process as in perception (you *see* that X is an apple) or supporting you while you work through an explicit inference ("White rump, short tail, pointy wings, swooping low: X must be a violet-crowned swallow"), is primarily an instrument developed to help its owner survive.[10] Given its very limited storage capacity, relative to the flood of raw data surging down the sensory channels when you're awake,[11] survival means fitting the stream of input into the available storage in such a way that it can be recovered when needed. The storage consists of a very large number of templates (a.k.a. patterns), in both the short-term and the long-term memory, plus a dating system and a cross-indexed filing system (using, as far as one can tell, metadata as relevance codes and a set of search heuristics). The templates are loosely grouped in the memory banks by knowledge management tasks and also arranged in a hierarchy of urgency that keeps the ones required for disaster avoidance, e.g., in traffic, family emergencies, or job interviews, near the top of the pile (metaphorically speaking).

The degree of compression here is impressive. Take the simplest way in which the brain compresses information—by using repetition indicators in short-term memory. If storing a 3D picture of this room in its entirety takes about 50Mb of information in a moderately high resolution hologram, storing the information that it has remained unchanged in most respects takes only 5 bits, and even if we note some small changes that have occurred and the duration of the basic sameness, we manage a reduction ratio of 1000 to 1. By using the short-term memory in this way, we can store very detailed patterns since we trash them, more or less completely, as we move to new surroundings and circumstances. The patterns that graduate to the big time circuit, i.e., the long-term memory, take up storage space that is at a much greater premium and much harder to reuse, but its payoff is also vastly greater. Patterns in long-term memory are linked to gatekeeper pattern recognition

[8] Some time ago, I provided an attempt to apply this kind of analysis to all the key concepts in the philosophy of science ("The psycho-logical foundations of modern science." *The New Metaphysical Foundations of Modern Science,* 1994. Noetic Sciences Institute, pp. 47–79.)

[9] I'm following the general geek' principle of good practice' here, which seems to be: If you can't find an algorithm, invent an acronym.

[10] Tony Blair, in his comments, rightly asked me to explain more clearly what I am saying about arguments by contrast with inferences, and this sentence is one of several in which I indicate my wish to support the view that the domain of inference (including probative inference) is vastly greater than that of arguments/reasoning and that logic should cover both without confusing them. The question "What made you think that?" so common in dialectic, covers both, and pushes for disclosure of cues as well as reasons, so that logic (the referee) can bring the rules to bear (e.g., locating the burden of proof and requiring a real proof. I'm just pushing to make sure 'real proof 'can include sound probative inference and its explicit formulation, including, for example, ceteris paribus clauses, but not an unlimited license to use them. A slogan for this effort might be 'Enthymeme, thy goal is deductive, but thy soul is probative."

[11] I've seen estimates that the brain might be only able to store three days' worth of sensory information if it had to store it all without shortcuts like the use of patterns. There is also a need to store internally generated information from e.g., reflection, inference, and speculation. Even if this is off by an order of magnitude or two, it remains clear that a categorical imperative of survival is brutal data compression in the nervous system.

routines in the sensory receptors and the front end processors of the main brain, and what they do is to cut a swathe through *all future input* in other words eliminating a potentially infinite amount of data that would otherwise have to be either stored or considered for storage. The pattern recognition process identifies new data that matches the pattern (roughly is good enough) as redundant since already stored, hence can be ignored. This is the great strength of even roughly correct patterns, and this is why the human skills at configural recognition are so highly developed, immensely more sensitive and accurate than the simple verbal forms we can express in our languages, even when they are enhanced by technical vocabularies. We recognize our friends and families although we are incapable of giving a description of them that would enable another to pick them out in a crowd.

So the storage of a pattern in the way gases behave, e.g., the general gas law, even though it's a very rough approximation to the truth, is all that we need for vast stretches of action scenarios, and takes up a million times less space than the raw data from a million more observations. The simplicity and accuracy of the patterns that the neo-positivists thought characterized natural laws adds a minor multiplier; rough generalizations are the icebreakers, exact laws mere ice cube makers, generating tidy tidbits of the good stuff.[12] Of course, such nice—although almost non-existent—laws *also* made deduction (or induction) of the explanandum possible, but by loosening that connection to that of probative inference from rough generalizations, we get an extremely effective result, with the advantage of not requiring false premises, viz., precise universal generalizations.

For our purposes here, which is constructing a block or two of the foundations of CT/IL, the main knowledge management tasks are things like: (i) finding the right name for some entity X that you're looking at, e.g., in order to communicate to someone about it, or to see if it's on your shopping list (in writing or memory), or on your list of dangers to run from or scream about; (ii) finding the right metaphor to use in explaining X to a particular listener; (iii) finding the right explanation of X in a scientific context; (iv) finding the right cause of X, in a fast practical search, for example of why your car has stopped running in the middle of a freeway; (v) finding the right rhyme for X in a poem you're writing; (vi) finding the right counter-example to an alleged generalization or definition, etc.

All these *finding* activities presuppose a stock of patterns in which to search, albeit a stock with an open texture, meaning that we can add to it by creative combinations or cross-references. Given that we have such a stock, the finding activity is a pattern-recognition one, usually handicapped by the limitations of language. Much of critical thinking is devoted to trying to express the considerations that underlie a clear feeling that an argument is invalid, a definition incorrect, or a generalization overreaching. This is the task of converting tacit knowledge to explicit knowledge, and we do it in the interest of validation, because we know that what seems like tacit knowledge can be inter-judge inconsistent, and hence is sometimes false intuition.

---

[12] When I first proposed this thesis (in "The Key Property of Physical Laws—Inaccuracy" in *Philosophical Studies* c. 1955, reprinted in Feigl & Maxwell (eds., 1961) *Current Issues in the Philosophy of Science*. New York: Holt, pp. 91–101), I was treated as someone who should be treated kindly, as one treats those having a slight mental disorder, by my good friends amongst the neo-positivists, e.g., Herbert Feigl and Peter Hempel. After all, everyone knew that *laws of nature* could not be inaccurate.

Probative inference fits into this account as, at times, a kind of promissory note for an inference that seems beyond reasonable doubt (BRD), although not meeting the standards of explicit deduction or induction. The expert birdwatcher identifying the violet-green swallow knows BRD that the identification is correct in an instant. When challenged for reasons, s/he reconstructs the criteria that influenced the call from memory. But there are times when BRD is present and the language can't carry the load, even for Gregory House; every birdwatcher and diagnostician knows of many such cases in their own experience, and every case of face processing by the man in the street also qualifies.[13] And, at the other end of the spectrum, there is the smoking/lung cancer kind of case, where all the cards are on the table, in words, and the case can be made bulletproof, given considerable time. Deduction and induction are of course the latter kind of case; probative inference has the middle ground, often but not always verbally reconstructable.
In terms of symbolic representation of probative inference, it's sometimes acceptable to portray them in the usual way. In the case of criterial inference this would be:

> This case has the properties C1, C2, C3,…. (where C are some of the criteria for X)
> ∴ This is an X

(Sometimes the C's would be cases, and the conclusion would be that X was the cause of the phenomena described in the cases; and there are other cases of probative inference.)
But it's sometimes desirable to be able to indicate that the inference is intended to be simply probative, so I propose that we indicate the special nature of a probative inference by using a double line under the premises. Then the above would be replaced by:

> This case has the properties C1, C2, C3,…. (where C are some of the criteria for X)
> ∴ This is an X

and in horizontal format would be represented as: C1, C2,…—>> X.

If the list of C's includes all the criteria for X, in the cases where they are all known, and known to be all, this would of course be a deduction.

In general, probative inferences can be challenged in the same way as claimed deductions, definitions, or generalizations, by citing counter-examples. For example, someone might say of the violet-green swallow case: "But there are two eagles that have those characteristics!" to which the response might be of the usual reformulation species: "Oh, well of course I'm taking overall size, and geographic/topographic location, and time of year into account as well."

So the suggestion here is that pattern recognition—claimed, challenged, refined—is the underlying process in much probative inference, whether of the taxonomic or causal or other sub-species. A corollary of this view is unwelcome to those of us with lingering dreams of logical calculi: the criteria of merit for probative inferences are mainly

[13] Though we are usually over the line between perception and inference with face recognition, we still can sometimes reconstruct the perceptual process, and a related point can be made for what we might call probative perception, i.e., conclusions can be achieved BRD without the help of inductive/deductive inference. (John Pollock's examples begin with cases of perception, not inference.)

the extent of the match between two patterns—the one we see now, and the template in memory with attached evidence of validity. We do have pretty good pattern-matching computer programs now, not just for fingerprints but for voices and faces; but the catch is that they are only rarely intelligible to humans when converted to common language. Logicians are even less fond than mathematicians of the idea that the validity of arguments is best left to the machines to judge; but that may be the best we can do in many cases. In others, the computers will point us to something that 'makes sense' at the level of the ordinary language of argumentation. One problem is that it's clear that training clearly improves the ability to identify good probative inferences, which lends too much credence to the idea that wine experts and art experts exist whose judgment of merit is any better than the lay person's, when in fact there are no grounds for thinking this at all. Of course, their judgments of authenticity (in art) or origin (in oenology) is testable but there's no way to bridge from such objective characteristics to merit—in these cases. In others, e.g., judging the merit of a redwood tree for cutting building lumber, merit is indeed a matter where skilled judgment is possible and provable.

I'd now like to consider a special variety of taxonomic probative inference—those with an evaluative conclusion. For example, in grading a student's answer to a short-essay question, we might reason from the presence of: (i) an original approach, (ii) that is clearly stated, and (iii) well argued, (iv) covering all that was asked, to the conclusion that it deserves an A grade in this context. Or that a Pilot G2 rollerball pen is an excellent model of that kind of writing instrument because of (i) its impeccably smooth delivery of (ii) an unbroken line of (iii) more than a mile's length per cartridge, (iv) even when the writing angle is as high as 30º, using a (v) highly stable ink, (vi) with excellent monochromatic hue, at (vii) the same price as other available choices; to the conclusion that it is a good balllpen.

Superficial inspection makes it clear that these are not deductions—other considerations about the alternatives need to be in place, and may be considered part of the obvious context hence omitted from the explicit consideration. In other words, a typical probative inference is useful just because it avoids getting into the almost endless quest for deductively adequate premises—or even into inductively high-probability premises—or the empty of alternative of turning it into a tautology by making it into a claim about an ideal entity. Instead, one can focus on the factors that account for all the variability, i.e., the significant relevant considerations, and settle for probative inference. We leave out reference to the *context* of the evaluative or taxonomic discussion, although it provides us with the *tacit* standards for merit or identification (the locale, season, and general size in the swallow case) because they are part of the common fixed background for participants in the communication, hence it's redundant to include them in the normal context of discussion. Incorporating this omission is part of the difference between the requirements on tools for practical reasoning and those for mathematical proof, e.g., between deduction and probative inference.

Note that we still critique these probative arguments by counter-exampling *within* that context. For example, the essay-grading illustration is clearly flawed by its failure to exclude plagiarism, currently a serious consideration if computers are allowed in exams where wi-fi is available, as it is on most college and many school campuses; and the rollerball example is flawed by the failure to cover leaking when carried, especially in hot weather or on planes, although those counter-examples stray close to non-standard contexts.

A point of particular interest about these evaluative examples of probative inference is that the criteria set for 'good X' or 'A-grade X' etc., are not derived in any simple way from the criteria for X. You need to know something about Xs and how they differ in ways that matter to users or those affected by the use of these things, in order to know how to tell good and bad ones apart. In other words, to define 'good apple,' just as to define 'apple' in the real world, you need to know a good deal about the criteria for each of them, and you can't get the first set from the second set. The dictionary is a useful tool, but like a pocket knife, it's not the full toolkit for a woodworker; and our knowledge of our native tongue covers the full toolkit, to use which requires that we know very well that a blade won't do the work of chisels and gouges.

In the version of this paper for publication I hope to include some real cases in detail from the first version of this paper, one of which provides the criteria for identifying a good 'chef's knife.' If you ask someone to explain to you exactly what a chef's knife ***is*** they can do this by talking about blade length, function, handle, and so on; but that won't tell you much about what makes a *good* chef's knife. Only someone with some knowledge of cutlery technology can do that, a group that mainly includes chefs who are interested in their tools as well as in cooking, kitchen equipment geeks, knife collectors, and the best knife designers, plus a few commercial kitchen equipment sales staff, though most of the latter group will not get it right.

Note that I am not talking about the highly debatable issues such as those surrounding a question like, Who makes the *best* chef's knife (or even, Is the Shun better than the Wusthof Icon), or Must a good chef's knife be made of steel; or of stainless steel; or can it be stellite, or ceramic? I'm just talking about what the dimensions of merit are, the criteria or properties on which performance has to be good in order that the entity qualify as a good chef's knife—and some crude idea of their relative weights, e.g., on a 4-way scale of Essential, Important, Marginally Important, Irrelevant. And some very crude idea of what minimum standards of performance are required for merit on those dimensions, e.g., that a chef's knife should hold its edge, perhaps with the help of a steel but not a sharpener, for at least one day of normal use by a chef. People don't think seriously about such matters unless they have an unusual vocational concern or other motivation such as collecting.

Even though all of us have eaten grapes for many years, and probably bought them most of that time, our knowledge of what counts as a good grape for eating is mostly tacit—but recoverable. There are at least a dozen criteria of merit for good table grapes, though you'd have trouble listing these if asked for them. But I could elicit these from you quite easily, the Socratic legerdemain, if I put a carefully selected dozen samples of bad grapes in front of you and asked you to say why each of them was unacceptable. So you know what makes a good table grape, though you probably don't know how to state what you know without cueing. In Ryle's terms, quoted with approval by Polanyi, you know *how* to identify them, but not *what* it is that distinguishes them. You have the pattern of merit in your head, but not the formula. So you are in a position to say with confidence: "Those grapes look very good" but it's a perception, not an inference. You might say the same about a knife in an illustrated catalogue, or even one with just the specifications.

The cutlery expert, on the other hand, will look at the actual knife rather carefully, test a few things about the knife, use a loupe to look at the grind on the edge, and can make a probative inference to the conclusion that it's *really* a good chef's knife. The premises of that inference are the facts about the knife, and the conclusion is about its merit, so it's an inference from empirical premises to an evaluative conclusion. The inference license is the presence of enough items from the list of criteria of merit to establish merit probatively, and since that list is what defines merit, it doesn't count as an extra premise. My take on Searle's example of the same species (promising behaviour implies obligation) would be that his premises turn out to require a string of conditions that would be hard to completely specify, just like the list of defining criteria in examples I've used here, and hence that the inference is in fact probative, although he claims it is deductive. As Tony Blair pointed out to me, he uses the language of prima facie and defeaters, so he might be receptive to this modification of his view.[14]

What about the classical objections to such inferences? On my view, Hume was right—you can't (strictly speaking) *deduce* evaluative conclusions from non-evaluative premises; but he was wrong in thinking that this means you can't *validly infer* them, i.e., probatively, which is all that should matter in any realistic logic. (That is, I won't concede the territory of validity to deduction, any more than the Oxford dictionaries do.) And Moore was right that any classic definition of 'good' (i.e., one offering a short synonymous translation) is questionable; since (i) 'good' can't be classically defined out of context, it can only have its meaning explained in the metalanguage (in terms of its modifier function in the process of evaluation): but (ii) "good X" *can* be criterially 'defined' in context (although "explained" is a better term), but that kind of definition is not threatened by the Moorean question, because it already concedes the (so-called 'logical') possibility of incompleteness by ending with an ellipsis. So Moore was wrong in thinking that his question refutes naturalistic approaches to ethics.

Hence, if we abandon the notion that classical definitions are the paradigm for explaining meaning, we can make informal logic a much better account of critical thinking in general, and evaluative reasoning in particular—and remove some alleged logical obstacles to naturalistic ethics. And, not so incidentally, do a better job of teaching critical thinking, i.e., one that will be more useful, especially because applicable in a wider range of contexts.

Finally, back to the pattern recognition theme for a wrap-up. Just as (i) identifying an X is a pattern recognition task, hinted at but not exhaustively analyzed by a criterial definition, so too is (ii) explaining X, or (iii) finding the cause of X, or (iv) classifying X, or (v) finding a way to program the task X, or (vi) fix the engine failure X, etc. Of course, success in these tasks requires not only the great pattern recognition skills your brain provides, it also requires a great repertoire of patterns to search, a task that early education and liberal arts education contribute to as well as specialized education e.g., a knowledge of the many ways a constant velocity joint can be constructed, and experience (e.g., in distinguishing the sounds of various mechanical flaws in operational but defective engines). There are ways to improve education in that direction, but our task here is just to focus on probative inference as the mechanism for pattern recognition. In that direction, the suggestion made here is that this part of the process is driven by the General Elimina-

[14] As a footnote to an old issue, perhaps one might say that Searles' argument is *contextually deductive* in what I take to be Sherlock Holmes' sense of 'deductive,' where there is some reference to 'this situation' in the required premises.

tion Method. The details of this vary amongst the six examples given, but let's take the hunt for causes as a typical case.

The GEM approach here requires the following premises:

1. Every event of the present type (X) has a cause (or causes). (True for almost all macro events.)
2. A putatively exhaustive list of the possible causes of an event of type X includes: C1, C2, …Cn
3. Each Ci in this list brings about X via a chain of intervening causes, Ci1, Ci2, … These are referred to as the *modus operandi* of the cause Ci. (There may be cases of direct action where Cij is zero.)
4. The facts of the case show that only Cp and Cq of the Ci were present prior to the occurrence of X.
5. The facts also show that the modus operandi of Cp but not Cq were all present.

   Hence, we can infer that Cp was the cause of X on this occasion.

This inference will be probative, with a degree of certainty that depends on that of the premises, and that level often exceeds all reasonable doubt. Whole sciences such as epidemiology, forensic medicine, criminalistics, and geology, are built on using GEM, and they are notably short on predictions, long on explanations. Support for probative inferences, when attacked, can come (at least) from explicating tacit knowledge (e.g., by expanding the lists of causes or MOs) or from holistic/gestalt arguments using metaphors or analogies to the pattern of the inference, or from appealing to (possibly rough) generalizations. This is (one aspect of) how the process of pattern recognition interfaces with explicit logic. Note that in some of these cases we will be dealing with approximations and other types of vagueness, and these cases need further discussion.[15]

I think we may need to create a new category, half way between tacit knowledge, which Polanyi *defines* as incapable of conversion to explicit knowledge, and the latter. We could call it latent or implicit inference. This will give us three types of probative inference, according to the degree of explicitness of the premises: (i) fully explicit—the smoking and lung cancer case; (ii) convertible by interrogation or challenge—the ornithologist case and many forensic cases e.g., the radiologist; (iii) the immutably tacit case, where the inferrer can't identify the cues but is confident of the inference and has solid evidence of past successes or concurrent corroboration to support it. The third case is still not perception, because there is no recognition involved and the inferrer states or agrees that an inference is being made. (Of course, s/he may *say* "It *looks* to me like a case where there's a real risk of suicide and we should put this patient on 24-hour watch" but that's a metaphorical use of 'looks.')

There are also large areas of applied science, e.g., policy analysis, city planning, internal medicine, that use a version of GEM to generate *predictions*. Here, premise 1' is: every intervention has an effect (or effects). The completeness of the premise corresponding to 2 above—the list of all possible effects, 2,'—is often shakier and hence the proba-

[15] Elia Zardini from the University of Edinburgh has been working on inexact knowledge and what he calls 'tolerant logics' that restrict transitivity of implication but permit a proof of consistency for the logic, in what looks like a promising start. (*The Reasoner* 4/09, p. 6)

tive inference to a prediction is made with less confidence. However, in this analogous process—it might be called the management or control version of GEM—we can still ensure at least some improvement of the probabilities of a desired outcome. We do this by the elimination of entries in the 2' list, not by inspection but by interventions that prevent the less desirable outcomes, in order to achieve (rather than reactively infer) the desired outcome.[16]

I hope that the GEM approach can thus be seen as a key process in the validation of one important family of probative inferences that is common in both empirical scientific work and in the many branches of evaluation (e.g., program evaluation and policy analysis) in science, engineering, technology, the humanities, and jurisprudence, and management or government. I am of course only calling attention to it, not inventing it, and labelling it in order to make the attention easier, and also perhaps to increase the ease of remembering the above modest (and tentative) clarification of it.

In conclusion, I hope that these rough ideas will encourage suggestions for improvement, and even perhaps some increased attention to the concept and use of probative inference in texts on both critical thinking and informal logic.

NOTE: This paper, given at OSSA in June, 2009, is the fourth version of a paper with this title, first circulated in late 2008, and has been much improved by comments on the first version by Tony Blair, Bob Ennis, and Mary Anne Warren; and on the second by those who commented on it at the Pacific APA meetings in April, 2009 (many thanks to Wanda Teays for facilitating that presentation). I'm especially indebted to Tony, whose detailed references to the existing literature were very helpful and are frequently used in this version. The third version, containing by then I think not one sentence from the first, went to John Woods for his comments at OSSA, June, 2009, in the light of which I am hopeful it will be worth developing a fifth version. I have (I think) improved it since sending the third version to John, by adding several hundred words, but will try to make clear where this has happened when presenting it. The predecessor of these efforts was an invited paper, "Probative Logic," at the 1st International Conference on Argumentation in Amsterdam, in 1986, that I think was published in its *Proceedings*. (I haven't looked at that since giving it, so this one may contradict it.) My continuing interest in the topic comes partly from an urge to get informal logic to portray ordinary reasoning more accurately, and partly from an effort to get the logic of the new discipline of evaluation clarified.

Link to commentary

[16] There's an interesting hybrid case in formative evaluation i.e., evaluation for improvement. Here the evaluator identifies weaknesses, and concludes that if these were remedied, the evaluand would be excellent. This evaluation involves a conditional prediction, conditional on a proactive response to the evaluation,