VIII. 3, Fall 1986
Informal Logic

A "Logical Audit" Scheme for Two-premise Arguments

WAYNE GRENNAN St. Mary's University

Introduction

In a paper entitled "A Logical Audit Scheme For Argument Evaluation" presented at the International Conference On Argumentation in June 1986 (to be published in the Conference Proceedings by Foris Publications) I presented a scheme for reliably arriving at an accurate judgment of the extent to which arguments prove their conclusions. I described the scheme as a refined version of one presented in my *Argument Evaluation* (University Press of America, 1984). The Conference paper presented the system as it applied to arguments containing only single premises. This paper will present a procedure for evaluating two-premise arguments.

The rating scheme described in the first paper (to be referred to as "Logical Audit") is based upon the use of a set of rating symbols and a diagram format for representing argument structure. The diagram format is a variant of that given by Michael Scriven in *Reasoning*. Each proposition is symbolized by a letter and each assertion is represented by a circle containing that letter, or several letters if it is a logically compound assertion. The circles containing the letters are connected together by arrows, with the arrow indicating the "direction" of inference. For example, the argument "The moon looks circular so it is a sphere." can be diagrammed as shown in Figure 1.

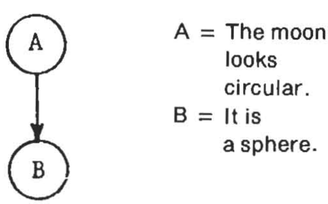


Figure 1

A Rating Scheme

Using this type of diagram format we can depict the logical structure of arguments with any amount of complexity. It can also be used for arguments cast in propositional logic by placing symbolic versions of the assertions in the circles.

We can also use the diagram as a place for recording our judgments about the quality of premises, inferences, and degree of proof.

The scheme being discussed here is a response to the problem of arriving at an accurate overall judgment of the quality of arguments. In earlier times when logic was considered to be concerned exclusively with inference quality there was thought to be no need for schemes such as this. More recently it has been recognized by writers of logic texts that this is not good enough. Scriven, for instance, exhorts the reader to "... make yourself give an overall grade. It's a cop-out not to. You must decide whether it does have force, and how much, for you." (*Reasoning*, page 45). To date, however, neither Scriven nor anyone else writing in this genre has developed a formal procedure for rating arguments in terms of the overall support they give their conclusions.

The scheme presented in "Logical Audit" uses four symbols that correspond to three sets of English expressions. There is a set for describing inference quality, a set for premise truth value judgments, and one for judgments of degree of proof. The single set of four rating symbols is used to report all of these ratings, so that each symbol has three uses. The scheme is set out in Table 1.

Table 1

Symbol	Probability Range	Premise Judgments	Inferential Support	Overall Support
0	Under ½ or ½	More likely false than true, or as likely.	Evidence against, or irrelevant.	None
+	½ to ⅔	Likely	Weak	Weak
+ +	⅔ to 5/6	Probable	Moderate	Moderate
+ + +	5/6 to 1	True	Strong	Strong

It will be obvious that there is some arbitrariness in choosing a four-term rating scheme, at least as regards having three to cover the probability range from ½ to 1. Constraints influencing this choice were: (1) a more fine-grained set would be clumsier to use and might be more precise than the accuracy of our judgments of quality can be; (2) a less fine-grained set, i.e. using only two symbols for the range from ½ to 1, would not do justice to our ability to discriminate in judging quality.

There is also arbitrariness in dividing the probability spectrum from ½ to 1 into three equal portions. In "Logical Audit" I argue that this arbitrariness is not vicious for two reasons. First, the meanings of the English rating expressions constrains us. For example, given that there are to be three intervals, it cannot be maintained that the boundary between "weak" and "moderate" must be set at ⅔ rather than 7/10, but the meanings of the two terms militate against setting it at ¾.

Secondly, the choice of interval boundaries, given the semantic constraint just mentioned, is not very critical. To show this, let us suppose that someone wished to regard the "++" increment covering the range from 0.7 to 0.9, so that "+" would cover the 0.5 to 0.7 increment and "+++" the 0.9 to 1.0 increment. Now let us also suppose that in some single-premise argument we agreed that the premise was "probable" and the inferential support was "moderate". If we use the product rule to ascertain the degree of proof and use the mean value for the "++" range in each case, the outcome using my definition is 0.56 (0.75 x 0.75). The outcome using the other definition is 0.64 (0.8 x 0.8). This may seem to be a significant difference, but at this point we need to recall that the goal of the rating scheme is to arrive at an accurate judgment of argument degree of proof expressed in a natural language such as English. This involves "translating" our verbal judgments of premise and inference quality into symbolic form, using the product rule to determine degree of proof expressed symbolically, then translating this symbol back into a verbal equivalent.

Applying this last step to the example, my scheme would commit me to judging the argument as providing weak support for its conclusion, since 0.56 falls in the "+" range. But note that 0.64 also falls within the "+" range as defined in the other scheme. Thus, using the other definitions still leads to the judgment that the argument provides weak support for its conclusion.

So long as we wish to divide the probability range from ½ to 1 into three increments and wish to use the English terms of appraisal that I use in the foregoing table, we will get agreement about degree of proof provided we start from the same inference and premise judgments, even if there is disagreement over the specific boundaries defining the probability increments. Thus, location of the boundary points is not critical.

Rating Single-Premise Arguments

Some of the value of expressing evaluation judgments using the special symbols can be shown by considering how they function in arguments having only one premise. With one premise the overall support for (degree-of-proof of) a conclusion is the product of the probability of the premise, p(P), and the probability of the conclusion given the truth of the premise, p(C/P). Now given the probability ranges assigned to each of the rating symbols and using this formula, we can express the overall support for the conclusion for various permutations of premise and inference ratings in a grid:

Premise Rating	Inference Rating 0	+	+ +	+ + +
0	0	0	0	0
+	0	0	0	+
+ +	0	0	+	+ +
+ + +	0	+	+ +	+ + +

The grid embodies the following rule: the argument rating equals the inference rating lowered by one value for each level that the premise is rated below "+ + +". For example, if an inference is rated "+ +" and a premise is also rated "+ +", we can find the argument rating by reasoning that the premise is rated one level below "+ + +" and reducing the inference rating one level yields a rating of "+".

Applying the procedure to the example given at the beginning, we would rate the premise A ("The moon looks circular") as true and enter "+ + +" beside the circle on the diagram. Turning to the inference, it is clear that A is relevant to B but would not come close to guaranteeing its truth, since objects that look circular from one vantage point can actually be ellipsoids, oblate spheroids, or even cylinders (seen end-on). On the other hand, A seems to be more than weak support for B. Let us rate the inference "+ +". Now we enter the premise and inference ratings on the diagram (Figure 2, below) and use the foregoing rationale to establish the overall rating of the argument. This will be "+ +", the same as the inference rating, since the premise is not rated below "+ + +". We now add this rating to the diagram (Figure 3).

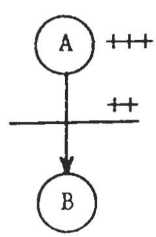


Figure 2

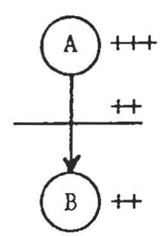


Figure 3

The above grid represents, in effect, a "recipe" for assigning an overall degree-of-proof rating to an argument whose premise and inference ratings have been determined. As such, this rating system (and similar ones) offers an important advantage over ratings approaches that operate without ratings symbols: it makes the step of judging degree-of-proof a *formal*, mechanical, step. Perceived weaknesses of an argument are reflected in premise and inference ratings and do not enter into making this judgment. In the example, once we decided on the premise and inference ratings, we obtained the overall rating independently of the argument. Thus, all arguments having these premise and inference ratings would receive the same overall rating.

With an informal, intuitive approach to an overall judgment of argument quality there is bound to be some unreliability arising when one attempts to combine premise and inference judgments to get an overall judgment. One factor that can be operative is a tendency to use some unarticulated personal algorithm other than the product rule.

In my experience such algorithms tend to yield over-positive judgments. For example, in "Logical Audit" I cite an instance of this observed in an Introductory Philosophy class. The students were asked to assess the degree of support for the claim that a dice shows a number under four when someone, who alone can see it and who tells the truth only 75% of the time, says that the number is under five. They were asked to describe the degree of support as "strong", "moderate", "weak", or "nil".

The students' situation can be conceived as one in which they must evaluate a three-assertion argument. In the diagram format assertion 'A' would be connected to 'B' which would in turn be connected to 'C'. 'A' stands for 'X says the dice shows under five', 'B' for 'The dice shows under five', and 'C' for 'The dice shows under four.' Using the rating scheme, A would be rated "+++", the inference from A to B would be rated "++" (p(B/A) = 0.75), and the inference from B to C would also be rated "++" since the probability that a dice shows under four, given it shows under five, is 0.75. Using the grid we find B should be rated "++", which results in C being rated "+". Thus, we would describe the support for C as "weak". The students' responses were typically over-positive: 50% said "moderate" and 21% even said "strong"!

Two-Premise Arguments

Whatever the virtues of this rating scheme in a logical audit of one-premise arguments, when inferences are made from more than one premise, the situation becomes more complex. The source of the complexity is the impossibility of being able to predict the probability of a conclusion, C, given two or more premises (P1, P2, etc.) when we know the inferential support that each premise provides individually. This problem arises from the probability relations among the elements of the argument. The formula expressing the relation is known as the Bayes "Inversion Theorem":

$$p(C/P1P2) = \frac{(p(C/P1) \times p(P2/P1))}{p(P2/P1)}$$

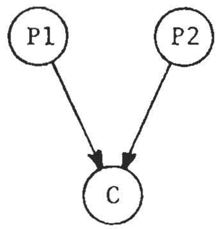
Note that in the formula the term p(P2/P1C) occurs, the probability that P2 is true given P1 and C. In commenting on the practical aspects of determining p(C/P1P2), William Kneale says the theorem "...is not very helpful. For any statistical information which enabled us to evaluate p(P2/P1C) would presumably enable us to evaluate p(C/P1P2) directly." (*Probability And Induction*, p. 129)

A more fundamental problem in determining p(C/P1P2) can be illustrated by adapting an example from Kneale. Suppose a painting is in a certain style of brushwork, and that this provides moderate inferential support for claiming it was a Rembrandt. Suppose also that it contains a certain pigment, and that this too provides moderate inferential support for claiming it was a Rembrandt. Regarding these two items of evidence as premises for the same conclusion we might be tempted to rate the inference as strong. But, Kneale says, "...we may conceivably have p(C/P1) and p(C/P2) both very high, but p(C/P1P2) very low or nil. For it may be that Rembrandt altered his style of brushwork during the course of his life and that when he used the distinctive style used in the picture the peculiar pigment found in the picture was not available. If we have reason to believe that this was so, we are compelled to say that the picture is not by Rembrandt." (pp. 128, 129)

The rating to be given the inference

when there are two or more premises, then, cannot be established reliably from the inferential support that each premise provides by itself. The inferential support of premises taken jointly must be judged on a case-by-case basis.

Unfortunately, matters become more complicated when we try to determine degree of proof. Given two premises, it might be thought that the probability of the conclusion was the product of the probabilities p(P1), p(P2), and p(C/P1P2). But consider this argument:



P1: Boston is more than 100 kms from New York.
P2: Boston is more than 1000 kms from New York.
C: Boston is more than 50 kms from New York.

Figure 4

In this argument p(C/P1P2) = 1, p(P1) = 1, and p(P2) = 0. Using the product rule we would have to conclude that p(C) = 0, and be committed to saying that the argument provides no support for its conclusion. But of course the conclusion is proved because P1 itself proves it.

The two arguments just discussed imply that there can be no definitive grid constructed for arguments with two premises, comparable to the one given for single-premise arguments. In what follows I attempt to provide strategies that can be used in place of such a grid, although the grid will have a role in the application of these strategies.

The second argument type can be dealt with by treating a two-premise argument as representing three distinct arguments: one containing P1 and C, one containing P2 and C, and one containing all three assertions. The principle on which this procedure is based is that the overall argument rating is the greater of the degree-of-proof ratings for these three arguments. Henceforth I shall refer to this as the "disjunctive" strategy. Applying this strategy to the last argument, we would end up rating the conclusion as proved because the argument from P1 to C proves its conclusion even though the other two arguments receive ratings of "0". This seems to be intuitively appropriate despite the fact that P2 is false. We might say that the argument has a logically benign deficiency.

Unfortunately, the disjunctive strategy does not work for the Kneale example or for other ones in which premises are discorroborating. In the Kneale example, taking an argument consisting of the premise "This painting was done using a brush technique Rembrandt once used" and the conclusion "This painting is a Rembrandt", we might rate the inference "++", the premise, "+++", and the degree-of-proof would be "++". Replacing the premise with "This painting contains such-and-such a pigment", we might rate the inference "+" and the premise "+++", so that the overall degree-of-proof would be "+". Evaluating the argument containing all three assertions, we would rate the degree-of-proof as "0" because the inference would be rated "0". Now following the disjunctive approach described above, we ought to conclude that the overall rating for the original argument is "++", the rating of the best argument. However, the appropriate rating is "0" since the premises, taken jointly, prove the conclusion is false.

Given these cases, the most reliable approach to rating two-premise arguments requires distinguishing between ones with discorroborating premises and the rest. When premises are discorroborating, we would rate the argument as a whole. When they are not, a vastly more common situation, we would apply the disjunctive strategy, then regard the correct rating as that of the highest-rated of the three arguments.

In what follows I develop a set of procedures reflecting this approach. To do so, it is necessary to adopt a set of symbols for the various ratings involved. The set is listed below.

Ri1: the rating of the inference from P1 to C
Ri2: the rating of the inference from P2 to C
Rij: the rating of the inference from P1 and P2 to C
Rp1: the rating of P1
Rp2: the rating of P2
Rc1: the proof rating of the argument from P1 to C
Rc2: the proof rating of the argument from P2 to C
Rcj: the proof rating of the argument from P1 and P2 to C
Ra: the final rating of the argument

The first five of the above values are independent variables. The next three, Rc1, Rc2, and Rcj are dependent variables. Ra denotes the correct rating for the argument. Values for Rc1 are determined from Rp1 and Ri1 using the grid presented earlier for one-premise arguments. Values for Rc2 are found the same way.

Rcj is determined using the product rule: p(C) = p(P1) x p(P2) x p(C/P1P2). The determination of Rcj can be done by using the grid for single-premise arguments in two steps. We can think in terms of calculating p(P1) x p(P2) first and then multiplying this product by p(C/P1P2). Now the grid values for Rc correspond to the product of two probabilities and it does not matter mathematically what these figures represent. Thus, given the ratings of two premises, the rating corresponding to the conjunction can be found from the grid. Then, that rating in conjunction with the inference rating can be used to get Rcj. As was stated before, the grid embodies a rule: for one-premise arguments the degree-of-proof rating equals the inference rating lowered by the number of levels below "+++" that the premise is rated. This can be generalized, given the possibility of combining any two ratings, as just described, to arguments with any number of premises: *Rcj equals Rij lowered by the total of the number of levels each premise is rated below "+++".*

Given these findings a procedure for arriving at the final rating, Ra, for a two-premise argument can be presented; see Table 2.

Table 2

STEP	QUERY	IF ANSWER IS YES	IF ANSWER IS NO:
1	Premises discorroborating?	Ra = Rcj	Go to step 2
2	Ri1 = "0"?	Ra = Rc2	Go to step 3
3	Ri2 = "0"?	Ra = Rc1	Go to step 4
4	Rp1 = "0"?	Ra = Rc2	Go to step 5
5	Rp2 = "0"?	Ra = Rc1	Ra = Higher of Rc1, Rc2, Rcj.

The preceding procedure applies when both premises are evidence for the conclusion. Often we are faced with a different situation: one premise is evidence for the conclusion but the other is what Toulmin calls a "warrant"—a premise intended to prove that the inference from the evidence to the conclusion is satisfactory.

Argument premises, according to Toulmin (*An Introduction To Reasoning*, ch 2) can either represent "grounds" (evidence) or "warrants" (principles that logically entitle us to infer a conclusion from grounds). Now warrants are by nature generalizations that might appear in a variety of arguments, so that an inference from a warrant alone to a particular conclusion would usually be rated "0". Yet warrants are relevant to the conclusion we wish to infer from our grounds, in the sense that a good warrant will result in Rij being greater than the rating for the inference from the ground to the conclusion.

As with the previous case in which neither premise counted as a warrant, there will be no guarantee that Rcj will equal or exceed Rc2 (if we call the warrant premise "P1"). Warrant and ground may be discorroborating, or the warrant may be false and redundant as it is in "Any place more than 1000 kms from another is more than 2000 kms from it. Boston is more than 100 kms from New York. Therefore, Boston is more than 50 kms. from New York." This is a proof even though the warrant is false and irrelevant, because Rc2 is "+++" even though Rcj would, by the product rule, be "0" because P1 (the warrant) is false. Of course, there is one difference from the two-grounds case: Rc1 will always be "0" since a warrant can only be evidence for the validity of the inference from P2 to C, which means that Ri1 is always rated "0". Ra, then, will be the higher of Rc2 or Rcj.

If the ground, P2, is irrelevant (Ri2 = 0) the argument rating will be "0". Since ex *hypothesi* no warrant can provide support for the inference from P2 to C in this case, the value of Rij will also be "0", which results in Rcj being "0". On the other hand, if Ri2 is not "0" but Rp2 = 0, both Rc2 = 0 and Rcj = 0. Thus, Ra = 0.

Given these results the procedure, when one premise is a warrant, is as shown in Table 3.

Table 3

STEP	QUERY	IF ANSWER IS YES	IF ANSWER IS NO
1	One premise a warrant?	Call it "P1" and go to step 2.	Not applicable.
2	Ri2 = 0 ?	Ra = 0	Go to step 3
3	Rp2 = 0 ?	Ra = 0	Go to step 4
4	Rp1 = 0 ?	Ra = Rc2	Ra = Higher of Rc2, Rcj

For convenience, the different procedures for the two types of cases can be combined into a single one, as in Table 4.

Table 4

STEP	QUERY	IF ANSWER IS YES	IF ANSWER IS NO
1	Premises discorroborating?	Ra = Rcj	Go to step 2
2	One Premise a warrant?	Call it "P1" and go to step 7.	Go to step 3
3	Ri1 = 0 ?	Ra = Rc2	Go to step 4
4	Ri2 = 0 ?	Ra = Rc1	Go to step 5
5	Rp1 = 0 ?	Ra = Rc2	Go to step 6
6	Rp2 = 0 ?	Ra = Rc1	Ra = Higher of Rc1, Rc2, Rcj
7	Ri2 = 0 ?	Ra = 0	Go to step 8
8	Rp2 = 0 ?	Ra = 0	Go to step 9
9	Rp1 = 0 ?	Ra = Rc2	Ra = Higher of Rc2 or Rcj.

Multi-Premise Arguments

In evaluating two-premise arguments it was found necessary to apply the disjunctive approach to arguments that do not have discorroborating premises. The same holds true for ones with three or more premises, although the task is more onerous. For two-premise arguments there were three arguments to rate. With three premises, there are not only Rcl, Rc2, Rc3, and Rcj to consider, there are also the three two-premise ones: P1 and P2 to C, P1 and P3 to C, and P2 and P3 to C. The value for Ra will be the rating of the higher-rated of these seven arguments. Although this sounds tedious, frequently it will not be. For example, we often find the premises acceptable, and when they are, the highest rating will be Rcj because the premises provide cumulative evidence. Rcj can be determined by the same formula used for two-premise arguments. When all premises are rated "+++" Rcj will be equivalent to Rij.

Conclusion

In this paper I have presented a rating scheme, and a procedure utilizing it, designed to enable us to arrive at an accurate personal estimate of the extent to which an argument proves its conclusion. It is important to note that the value for Ra that a person arrives at is supposed to be based on all the evidence (for and against) bearing on the premises that one has at one's disposal, plus all the knowledge one has about the extent to which the conclusion follows from the premises. Direct evidence for the conclusion is not supposed to play a part in the evaluation of the argument, as the object is to determine the extent to which the proffered premises support the conclusion. The object is not to establish the truth of the conclusion by whatever information one has. To do this would be to ignore the argument.

The concern for degree of proof rather than truth has an important bearing on the persuasiveness of arguments. Normally an argument is presented by an arguer on the assumption that the arguee does not accept the conclusion as true due to a lack of evidence in her/his possession. The arguer is attempting to remedy this deficiency. Sometimes, however, the arguee already had evidence for the conclusion that provides better suppport for it than the argument gives. In such a case the conclusion rating (Ra) that the arguee would give is lower than the rating that the conclusion itself would be given independent of the argument. In such cases the arguee might give Ra a rating of "++" or "+" but regard the argument as having no persuasive value for her/him, although someone else assigning the same value to Ra might find that it does have persuasive value for them.

Given the rating scheme and the procedure, it is intelligible to talk of the "correct rating" for premises, inferences, and arguments. The correct rating for a premise is the one that would be given by anyone in possession of all the significant evidence for and against the premise, who was capable of "weighing" that evidence accurately. The correct rating for an inference would be the one arrived at by someone possessing all the significant information pertinent to deciding the extent to which the truth of the premises would guarantee the truth of the conclusion. The goal in argument evaluation is to arrive at the correct rating of an argument using the appropriate information. But, as we have seen, even if we are able to arrive at the correct rating, we might not find the argument has persuasive value. However, it should be obvious that this cannot be the basis for a criticism of the argument evaluation system presented in this paper. I cannot elaborate here, but one might use the argument rating one arrives at (Ra), in conjunction with one's rating of the conclusion taken alone (Rc), to develop a quantified measure of the persuasiveness of any argument one encounters. The difference between these two ratings reflects persuasiveness: an argument has some persuasive value when Ra exceeds Rc.

Dr. Wayne Grennan, Department of Philosophy, St. Mary's University, Halifax, Nova Scotia B3H 3C3 □