Confirming Inexact Generalizations

Ernest W. Adams

University of California–Berkeley

1. The Formal Language

An inexact generalization like 'ravens are black' will be symbolized as a propositional function with free variables thus: 'Rx ⇒ Bx.' The *antecedent* 'Rx' and *consequent* 'Bx' will themselves be called *absolute* formulas, while the result of writing the non-boolean connective '⇒' between them is *conditional.* Absolute formulas are arbitrary first-order formulas and include the exact generalization '(x)(Rx → Bx)' and sentences with individual constants like 'Rc & Bc.' On the other hand the non-boolean conditional '⇒' can only occur as the main connective in a formula. We shall also need to consider formulas with more than one free variable such as 'xHy ⇒ xTy,' which might express 'if x is the husband of y then x is taller than y.' Though it is inessential, it will simplify things to work in 'n-languages' with a finite number of individual constants $c_1, \ldots, c_n$, which are interpreted as denoting the elements of the domains of the 'n-models' to be described below.

2. Degrees of Truth

Sentences like '(x)(Rx -> Bx)' and 'Rc & Bc' are defined to be true or false in a model M in the standard way, but the d*egree of truth* of a conditional formula like 'Rx ⇒ Bx' in M is defined to be the proportion of values of 'x' that satisfy 'Bx' in M out of all of those that satisfy 'Rx," and the degree of truth of 'xHy ⇒ xTy' is the proportion of values of 'x' and 'y' simultaneously that satisfy 'xTy' out of those that satisfy 'xHy.' These will be written Tr(Rx ⇒ Bx) and Tr(xHy ⇒ xTy), and Tr((x)(Rx → Bx)) is stipulated to equal to 1 or 0 according as '(x)(Rx → Bx)' is true or false in M. It is easily seen that on this definition Tr(Rx ⇒ Bx) is not necessarily equal to Tr(-Bx ⇒ -Rx)). The definition of Tr must be generalized in the case of infinite models, where proportions are not defined. Previous papers (Adams 1974, 1986, Carlstrom 1975, and Adams and Carlstrom 1979) describe generalizations and investigate the logic of this semantics, which is closely related to work of Douglas Hoover (1978, 1982). I will side-step the problem here by confining attention to *n-models* whose domains can be assumed to be just $\{1, 2, \ldots, n\}$. Assuming this, the degree of truth of an absolute formula $\varphi(x)$ with free variable 'x' is the mean of the truth values of its 'n-model instances' $\varphi(c_i)$ that result when 'x' is replaced by 'c_i' in $\varphi(x)$.

The crucial fact about the degree-of-truth measures just described is that they are *independent and symmetric probability functions*. Thus, Tr satisfies the Kolmogorov

PSA 1988, Volume 1, pp. 10-16

https://doi.org/10.1086/psaprocbienmeetp.1988.1.192963 Published online by Cambridge University Press

axioms for absolute formulas, and it is a conditional probability function in application to conditional ones. For instance, we can write Tr(Rx ⇒ Bx) = Tr(Bx/Rx) = Tr(Rx & Bx)/Tr(Rx) (if Tr(Rx) = 0 we arbitrarily stipulate that Tr(Rx ⇒ Bx) = 1). Tr is independent in the sense that if φ and ψ are absolute and have no free variables in common then $Tr(\varphi \& \psi) = Tr(\varphi) x Tr(\psi)$, and it is symmetric in the sense that if φ and ψ are alphabetic variants of one another then $Tr(\varphi) = Tr(\psi)$. Note that this is essentially *exchangeability at the level of variables*, as against exchangeability at the level of individual constants. The former is logically valid because alphabetic variants express the same propositions, while the latter is factual if it is valid at all.

3. Probabilities

The fact that degree-of-truth functions satisfy the probability axioms led both Lukasiewicz (1913) and Hoover to regard them as 'real probabilities.' However, the independence of these functions would not allow us to account for confirmation in terms of them, and we need to consider a more general concept of probability to do that. We can utilize Hoover's (1982, Theorem 6.3) version of de Finetti's theorem for this, which states that any symmetric probability function on a language is a mixture of independent and symmetric probability functions on the language; i.e., it is a mixture of degree-of-truth functions (Hoover excludes classical quantifiers from his languages, but that is inessential in the finite cases we are considering). I will confine attention to mixtures of degree-of-truth functions in n-models. Given this and the fact that the degree of truth of an absolute formula $\varphi(x)$ with free variable 'x' is the mean of the degrees of truth of its n-model instances $\varphi(c_i)$, we get the following dual representation of the probability of $\varphi(x)$:

(1) $$Pr(\varphi) = \sum_{j=1}^{N} Pr(M_j) x Tr_{M_j}(\varphi) = \frac{1}{n} \sum_{i=1}^{n} Pr(\varphi(c_i))$$

where $Pr(\varphi)$ is the probability of φ, $Pr(\varphi(c_i))$ is that of its n-model instance $\varphi(c_i)$, $Pr(M_j)$ is the 'weight' of n-model M_j, Tr_{M_j} is the degree-of-truth function for M_j, and we assume that there are N of these models. *The probability of an absolute formula is equal to its expected degree of truth in n-models, and to the mean of the probabilities of its n-model instances*. The first clause tells us that our concept of probability is a generalization of the ordinary concept of probability applying to sentences, whose probabilities equal their expectations of being true. We will assume that any distribution n-model probabilities $Pr(M_1), \ldots, Pr(M_N)$ is possible. That we do not assume exchangeability at the level of individual constants means that we do not assume that isomorphic n-models are equiprobable or that n-model instances are, which is one of the differences between our present approach to confirmation and those of Carnap (1950), Hintikka (1966), and many others. Note, though, that (1) implies that when the n-model instances are equiprobable they are equal in probability to the inexact generalizations of which they are instances, and in a sense inexact generalizations can be looked upon as 'average instance propositions' from the probabilistic point of view.

Another important property of probabilities leads in the direction of confirmation. Let φ be any absolute formula, and let $\boldsymbol{n}$ be any formula formed from conjunctions and disjunctions of alphabetic variants of φ, having no free variables in common with each other or with φ. Then it follows from Tchebycheff's Inequality (Hardy, Littlewood, and Polya 1952, Theorem 43) that:

(2) $Pr(\varphi/\boldsymbol{n}) \geq Pr(\varphi)$.

For example, φ might be 'Bx' and $\boldsymbol{n}$ might be '(By v (Bz & Bw),' in which case (2) implies that Pr(Bx/By v (Bz & Bw)) ≥ Pr(Bx). It is important to stress that we cannot omit

https://doi.org/10.1086/psaprocbienmeetp.1988.1.192963 Published online by Cambridge University Press

the qualification that the alphabetic variants forming n cannot have free variables in common with each other or with φ. For instance, neither $Pr(xTy/yTx) \geq Pr(xTy)$ nor $Pr(xTy/zTw \& wTz) \geq Pr(xTy)$ is valid in general. Also, free variables cannot be replaced by individual constants in the inequality. For instance, while necessarily $Pr(Bx/By) \geq Pr(Bx)$, it can happen that $Pr(Bx/Bc_i) < Pr(Bx)$ for particular c_i (note that this cannot happen with the exact generalization '(x)Bx'; that is because it logically entails Bc_i while the inexact generalization does not). This is important as regards confirmation, since evidence confirming a generalization typically has the form of *particular* data or instances and not of other generalizations. We can, however, relate conditional probabilities like P(Bx/By) to 'average' particular instance conditional probabilities as follows.

Let φ, ψ, and $n(y)$ be absolute formulas such that 'y' does not occur in φ or ψ. Then it follows from (1) that:

$$(3)\quad Pr(\varphi/\psi \,\&\, n(y)) = \frac{\sum_{i=1}^{n} Pr(\psi \,\&\, n(c_i))Pr(\varphi/\psi \,\&\, n(c_i))}{\sum_{i=1}^{n} Pr(\psi \,\&\, n(c_i))}$$

This tells us that $Pr(\varphi/\psi \& n(y))$ is a weighted average of the 'n-model instance probabilities' $Pr(\varphi/\psi \& n(c_i))$. One particular case is that in which φ is 'Bx,' $n(y)$ is 'By' and ψ is a vacuous tautology, where (3) reduces to:

$$(3a)\quad Pr(Bx/By) = \frac{\sum_{i=1}^{n} Pr(Bc_i)Pr(Bx/Bc_i)}{\sum_{i=1}^{n} Pr(Bc_i)}$$

Assuming the equiprobability of n-model instances it would follow that $Pr(Bx/By) = Pr(Bx/Bc_i)$ and therefore by (2) that $Pr(Bx/Bc_i) \geq Pr(Bx)$, which would make it appear that instances of absolute inexact generalizations 'weakly confirm' them. Not assuming instance equiprobability, we can only say that this must hold 'on the average,' though it can fail in particular cases. But we are interested primarily in the confirmation of conditional generalizations, and we will now see that not only may particular instance disconfirm them but all of them may do so. We will focus primarily on 'ravens are black,' symbolized as '$Rx \Rightarrow Bx$,' and we note the following special case of (3) which is relevant:

$$(3b)\quad Pr(Bx/Rx \,\&\, Ry \,\&\, By) = \frac{\sum_{i=1}^{n} Pr(Rx \,\&\, Rc_i \,\&\, Bc_i)Pr(Bx/Rx \,\&\, Rc_i \,\&\, Bc_i)}{\sum_{i=1}^{n} Pr(Rx \,\&\, Rc_i \,\&\, Bc_i)}$$

$Pr(Bx/Rx \& Ry \& By)$ is therefore a weighted average of n-model instance probabilities $Pr(Bx/Rx \& Rc_i \& Bc_i)$. The latter relate to Bayesian confirmation of '$Rx \Rightarrow Bx$.'

4. Confirming 'Ravens are Black'

Let us say that evidence weakly confirms '$Rx \Rightarrow Bx$' if the probability attaching to it after n is learned is at least as great as it was before. We will focus on that, and assume

https://doi.org/10.1086/psaprocbienmeetp.1988.1.192963 Published online by Cambridge University Press

that evidence must be expressed by sentences. We will also assume that the posterior probability of an absolute formula φ after learning n is given by the conditional probability $Pr(\varphi/n)$ (Bayesian confirmation, cf. Rosenkrantz 1977, Chapter 2), and if this is so and $Pr(Rx \Rightarrow Bx) = Pr(Bx/Rx) = Pr(Rx \& Bx)/Pr(Rx)$, it follows that the posterior probability of 'Rx $\Rightarrow$ Bx' after learning n must equal $Pr(Bx/Rx \& n)$. Therefore we may say that n *weakly confirms 'Rx $\Rightarrow$ Bx' if $Pr(Bx/Rx \& n) \geq Pr(Bx/Rx)$.* I will focus on *positive instances* which are data to the effect that particular ravens are black, and I will make the default assumption that they can be expressed by formulas of the form $Rc_i \& Bc_i$. This is questionable in view of the fact the constants c_i should properly symbolize names, but establishing that a raven is black is not usually to establish that some named raven is black. This matter deserves more attention that it has yet been given, but it cannot be pursued here (note the cryptic remarks about instances being symbolized by 'new constants,' in Carnap 1950, p. 572).

Now I want to show that not only can individual positive instances disconfirm 'Rx $\Rightarrow$ Bx', but all of them can. This can be seen intuitively. Suppose there were only two possible worlds, in the first of which 1% of all things were ravens and all of them were black, and in the second of which everything was a raven but only half of them were black. Suppose also that *a priori* it was 99% probable that we were in the first world. Then the discovery that any particular c_i was a black raven would lower the expected degree of truth of 'ravens are black' by making it much more likely that we were in the second world, in which only 50% of ravens are black; i.e., in this situation $Pr(Bx/Rx \& Rc_i \& Bc_i)$ would be *less* than $Pr(Bx/Rx)$ for all c_i.

Given our default assumption that positive instances of the form $Rc_i \& Bc_i$ are representative of positive instances of 'ravens are black' in general, it would seem that there is no logical guarantee that each positive instance will not disconfirm it. This should not be surprising in the light of Goodman's grue (Goodman 1955; this will be returned to briefly below), and it suggests that if we are to explain the intuitions that underlie commonly accepted inductive principles (cf. Russell 1912) we must appeal to Goodmanesque hypotheses about lawlikeness or natural kinds, or to related '*a priori* synthetic' assumptions about prior probabilities. I believe there is something right in this, but I want to end this part of the discussion by pointing out a kind of assumption which I suspect is commonly make in practice, but which has usually been overlooked in theory.

We noted earlier that $Pr(Bx/Rx \& Ry \& By)$ is a weighted average of the probabilities $Pr(Bx/Rx \& Rc_i \& Bc_i)$, all of which are less than $Pr(Bx/Rx)$ in our anomalous example; hence $Pr(Bx/Rx \& Ry \& By) < Pr(Bx/Rx)$ must have been the case in that example. Conversely, if we could describe special circumstances in which the inequality

(4) $Pr(Bx/Rx \& Ry \& By) \geq Pr(Bx/Rx)$

held it would follow that 'on the average' $Pr(Bx/Rx \& Rc_i \& Bc_i) \geq Pr(Bx/Rx)$ in those circumstances; i.e., on the average positive instances of 'Rx $\Rightarrow$ Bx' would have to weakly confirm it in those circumstances. The following pair of conditions would guarantee this:

(5) $Pr(Bx/Rx \& Ry \& By) \geq Pr(Bx/Rx \& Ry)$

(6) $Pr(Bx/Rx \& Ry) = Pr(Bx/Rx)$.

Clearly (6) fails in the anomalous example, and in fact it must fail in any circumstances in which (4) does. That is because (5) must hold under all circumstances, since it follows from a trivial generalization of (2).

But consider what (6) means. By (3) again, $Pr(Bx/Rx \& Ry)$ is an average posterior probability of 'Rx$\Rightarrow$ Bx' given data of the form Rc_i, which on our default assumption is the

https://doi.org/10.1086/psaprocbienmeetp.1988.1.192963 Published online by Cambridge University Press

average posterior probability of 'ravens are black' just given information to the effect that something is a raven. In the anomalous example this sort of information did lower the probability of 'ravens are black' by making it probable that we were in a world where a low proportion of ravens were black, but I would hazard that in normal circumstances in which conditional generalizations are tested the mere discovery that something was an *instance* of the generalization, without determining whether it was positive or negative, would not affect the generalization's probability. In other words, I hazard that while (6) may fail in anomalous circumstances, it holds in normal ones. I will call this the principle of the *Independence of Pure Instancehood* (IPI), and where it holds (4) must hold as well; i.e., when IPI holds positive instances of generalizations must weakly confirm them on the average.

I will conclude with very brief comments on applications, primarily stressing generalizations.

5. The Paradoxes and Other Applications

Let us begin with Hempel's Paradox (Hempel 1945), and focus on the confirmation of 'ravens are black' by sentences of the form 'Rc & Bc,' and by ones of the form '-Bc & -Rc' which would be positive instances of the contrapositive generalization 'non-black things are not ravens.' If the two generalizations were equivalent they should be equally confirmed (the "Equivalence Principle'), and if positive instances always confirmed then both instances should confirm both generalizations (Nicod's Principle). Standard Bayesian resolutions of this paradox reject or qualify Nicod's Principle (Suppes 1966, Rosenkrantz 1977) but accept the Equivalence Principle. My approach implicitly calls that into question also, by representing the generalizations involved as inexact rather than exact (they could also be regarded as inequivalent exact generalizations, for instance if they had Aristotelian existential import, but I will ignore that). However, it is more significant that IPI implies that the confirmation of 'Rx ⇒ Bx' by instances of form '-Bc_i & -Rc_i' is not only not always positive, it is necessarily non-positive on the average. That is because Pr(Bx/Rx) = Pr(Bx/Rx & -Ry) by IPI, and Pr(Bx/Rx & -Ry) ≥ Pr(Bx/Rx & -By & -Ry) by a non-trivial generalization of (2), hence Pr(Bx/Rx) ⇒ Pr(Bx/Rx & -By & -Ry). But the latter is the average posterior probability of 'Rx ⇒ Bx' given instances of form '-Bc_i & -Rc_i'.

Turning to the Goodman Paradox (Goodman 1955, pp. 74-5), consider 'emeralds are grue' interpreted as generalizing both over emeralds and over times at which they might be grue, where 'grue' is defined in the familiar way. This double generalization is appropriately symbolized as an open formula in a two-sorted logic, to which degrees of truth, probabilities, and the confirmation formulas given in the earlier sections all generalize directly. I will not go into details, but I would suggest that as things are, and unlike 'emeralds are green,' IPI would be unlikely to apply to 'emeralds are grue.' The reason is simple. As things are, persons would be likely to accept 'emeralds are grue' only if they thought that emeralds wouldn't exist after the time t at which greenness and grueness split apart. Assuming this, finding an emerald of whatever color shortly before t would disconfirm 'emeralds are grue' by making it likely that emeralds would exist after time t. This is in the spirit of Goodman's own solution to the paradox since it stresses the way certain kinds of evidence influence our beliefs *as things are*, rather than how 'in logic' they should influence them. Obviously, as in Goodman's approach, the probabilistic approach requires much more detailed study which cannot be entered into here. I will end by briefly citing applications involving two other sorts of generalizations.

One is to multiple-variable generalizations such as transitivity laws of the form 'if xRy and yRz then xRz' (e.g., 'if x can beat y and y can beat z then x can beat z'). These are expressible in the present formalism, and one interesting study has to do with relations between different expressions of seemingly the same generalization, such as between '(xRy & yRz) ⇒ xRz' and '(∃y)(xRy & yRz) ⇒ xRz.' It is easily seen that these are not necessarily equivalent in either degree of truth or probability, and the investigation of the factual conditions in which they are equal, at least approximately, leads to 'density' assumptions whose role in the theory of inexact ordering relations (Adams and Carlstrom 1979) is in some ways analogous to that of IPI.

The other application is to inference by analogy, as discussed in Carnap (1950 pp. 569-70). Here instead of generalization with respect to individuals it is with respect to predicates, and the natural approach is *via* second-order logic. There is no difficulty in carrying this out, provided that predicates are restricted in somewhat the way individuals were restricted in the n-languages considered here. Analogues of laws (1) - (3) can be derived straightforwardly, and plausible analogues of non-logical principles like IPI can be formulated. The special difficulty that arises in this case is to describe positive instances. The reader can get some idea of the problem by considering first-order absolute generalizations of the form 'Rx <-> Bx' and regarding 'R' and B' as names of individuals and 'x' as ranging over their predicates. We would normally regard formulas of the form 'Rc_i & Bc_i' or '-Rc_i & -Bc_i' as positive instances of this, but in fact they to not stand to it as positive instances of 'Rx ⇒ Bx' stand to that formula. This is related to the fact that Hempel's Consequence Principle (cf., Carnap 1950, p. 471) fails. The special circumstances in which it holds is a further subject for investigation.

References

Adams, E.W. (1975). "The Logic of 'Almost all'." *Journal of Philosophical Logic* 3: 3-17.

________. (1986). "Continuity and Idealizability of Approximate Generalizations." *Synthese* 67: 439-376.

Adams, E.W. and Carlstrom, I.R. (1979). "Representing Approximate Ordering and Equivalence Relations." *Journal of Mathematical Psychology* 19: 182-207.

Carlstrom, I.F. (1975). "Truth and Entailment for a Vague Quantifier." *Synthese* 30: 461-495.

Carnap, R. (1950). *The Logical Foundations of Probability*. Chicago: University of Chicago Press.

Goodman, N. (1955). *Fact, Fiction, and Forecast*. Cambridge: Harvard University Press.

Hardy, G.H., Littlewood, J.E. and Polya, G. (1952). *Inequalities* 2nd. edition. Cambridge: The University Press.

Hempel, C.G. (1945). "Studies in the Logic of Confirmation." *Mind* 54: 1-26 and 97-121.

Hintikka, J. (1966). "A Two-Dimensional Continuum of Inductive Methods." In *Aspects of Inductive Logic*. Edited by J. Hintikka and P. Suppes. Amsterdam: North-Holland Publishing Company, pp. 113-132.

Hoover, D.N. (1978). "Probability Logic." *Annals of Mathematical Logic* 14: 287-313.

_ _ _ _ _ _ _ _ _. (1982). "A Normal Form Theorem for L_w with Applications." *The Journal of Symbolic Logic* 47: 605-624.

Lukasiewicz, J. (1913). "*Die Logische Grundlagen der Wahrscheinlichkeits-rechnung*." Reprinted as "Logical Foundations of Probability Theory" in *Jan Lukasiewicz Selected Works*. Edited by L. Borkowski. Amsterdam: North-Holland.

Rosenkrantz, R.D. (1977). *Inference, Method and Decision*. Dordrecht: Reidel Publishing Company.

Russell, B. (1912). "On Induction." Chapter VI of *The Problems of Philosophy*. London: Oxford University Press.

Suppes, P. (1966). "A Bayesian Approach to the Paradoxes of Confirmation." In *Aspects of Inductive Logic*. Edited by J. Hintikka and P. Suppes. Amsterdam: North-Holland Publishing Company, pp. 198-207.

https://doi.org/10.1086/psaprocbienmeetp.1988.1.192963 Published online by Cambridge University Press