HAIM GAIFMAN

SUBJECTIVE PROBABILITY, NATURAL PREDICATES AND HEMPEL'S RAVENS

1. INTRODUCTION

The title well represents this paper's goals. I shall discuss certain basic issues pertaining to subjective probability and, in particular, the point at which the concept of natural predicates is necessary within the probabilistic framework. Hempel's well-known puzzle of ravens serves as a starting point and as a concrete example. I begin by describing in §2 four solutions that have been proposed. Two of these represent fundamental approaches that concern me most: the probabilistic standard solution and what I refer to as the natural-predicates solution. The first is essentially due to various investigators, among them Hempel himself. The second has been proposed by Quine in his 'Natural kinds'; it represents a general line rather than a single precise solution. Underlying it is some classification of properties (or, to remain safely on the linguistic level, of predicates) which derives from epistemic or pragmatic factors and is, at least prima facie, irreducible to distinctions in terms of logical structure. Goodman's concept of entrenchment belongs here as well (his paradox is taken up in §3 and §5). Of the other two, the one referred to as a "nearly-all"-solution is based on interpreting 'all' (in 'all ravens are black') as nearly all. An analysis shows that the valid part of this argument is reducible to the standard probabilistic solution. The remaining solution is based on a modal interpretation; it is shown to belong to the natural-predicates brand. Another modality argument turns out, upon analysis, to be false.

Though the issue of natural predicates is extremely complex, the naturalness of predicates is expressible in probabilistic terms and this can be done without going into the tangle of epistemic or pragmatic factors. Thus, the standard probabilistic solution (based on frequency considerations) and the natural-predicates solution that appear as incommensurably different views can, in fact, be combined and treated within a single framework. One can point out and even calculate the contribution of each factor to Hempel's puzzle: the frequency factor of the standard solution

Erkenntnis **14** (1979) 105–147. 0165–0106/79/0142–0105 $04.30

and natural-predicates factor. We shall see that, although the factor of naturalness does play a certain role in this situation, the puzzle cannot be explained by it, and the core of any true explanation must consist of the frequency considerations of the standard approach. The argument is given in §5, first in general qualitative form and then in a quantitative way involving the calculation of numerical values. As against this it is pointed out in §3 that Goodman's puzzle cannot be accounted for on standard probabilistic grounds. Its explanation must invoke some extra-logical distinctions of the natural-predicates variety; it exemplifies the general principle that our choice of a prior probability distribution is determined by extralogical factors. Thus, the verdict is that Quine has been wrong in assimilating Hempel's puzzle into Goodman's and in treating both via the concept of natural kinds. We have here essentially different riddles. This conclusion has been, probably, arrived at by various investigators who looked carefully into the problem, but, to my knowledge, this is the first time that the two approaches have been combined and the contribution of each factor precisely calculated.

In §3 the framework of subjective probability is presented and some inherent problems pointed out, in particular the question of a "good" prior. The measurement of the intuitive concept of confirmation is discussed and a certain function is proposed that is used later in §5. The way of presenting Goodman's puzzle within the probability framework is indicated. §4 contains certain observations concerning natural predicates and frequencies and the concluding §5 contains the final analysis of Hempel's puzzle. This analysis is intended also as an example of how the framework of subjective probability can be used in a philosophic investigation.

I could not resist certain observations that are marginal to the main issues but which, I thought, are worth making, and, to my knowledge, are original. In particular, the well-known argument purporting to show, by means of betting odds, that universal statements with infinitely many empirical instances should have probability zero is analysed in Note 4 and found to be erroneous. Note 5 contains a brief critique of Goodman's view and Note 10, conjointly with the relevant text of §3, indicates how Goodman's puzzle can be presented and clarified (though not solved) using the probability framework.

The paper is intended, among other things, as a final analysis of Hempel's puzzle. It is not, however, intended as a summing up of the considerable literature that this well known puzzle has evoked. In general, I omitted references to the four solutions discussed in §2 which, I hope, the authors will excuse.

2. THE PUZZLE AND SOME PROPOSED SOLUTIONS

'All men are mortal', 'all days are followed by nights', 'all ravens are black' – belief in the truth of these and similar generalizations has been explained in inductivist tradition as due to repeated observations of particular instances. The generalizations in question have the form 'All *A*'s are *B*'s' and Nicod proposed a rule by which every additional *A* that turns out to be a *B* adds further support. The rule presupposes, of course, that no counterexamples (*A*'s which are not *B*'s) have so far been discovered. It purports to explain how general statements get accepted. Thus, in order to establish the law that all ravens are black an investigator looks for ravens; every additional black one will serve as further confirming evidence. Hempel invented a lazy investigator who saves himself the trouble of looking for ravens simply by applying Nicod's rule to the generalization 'All non-black objects are not ravens'. Remarking non-black objects in his rooms and finding them, upon checking, to be non-ravens he obtains confirming evidence for this last general statement. Since, on logical grounds, the two statements are equivalent ('all *A*'s are *B*'s' is logically equivalent to 'all non-*B*'s are not *A*'s'), evidence which confirms one confirms to the same extent the other. Yet the confirmation of the law that all ravens are black by the checking of non-black objects in one's rooms is evidently unacceptable.[1]

A solution of the puzzle should explain why, of these two experimental procedures that accord equally with Nicod's rule, only the first is accepted. The difference should be derived from general principles of learning from experience. It is also desirable that the explanation throw some light on Nicod's rule.

The Standard (Probabilistic) Solution

A testing of an hypothesis consists, usually, in checking the truth of some consequence of it. Given that *a* is an *A*, '*a* is a *B*' is a consequence of 'all *A*'s are *B*'s'. Nicod's rule is therefore a special case of confirming by verification of a consequence. By "consequence" I mean any statement which, in view of existing knowledge, is implied by the hypothesis. The verification causes the hypothesis' probability to be multiplied by the inverse of the consequence's probability (the probability it had before verification). This is a general rule and a special case of Bayes' law, to be discussed in §3. Since this probability is, like any other probabilistic value, at most 1, its inverse is at least 1, implying that the verification cannot make the hypothesis' probability smaller. If, however, the consequence has already probability 1, or very near 1, practically no increase takes place; in as much as the verified statement is a foregone conclusion it is valueless as confirmatory evidence.

By finding a new black raven the first investigator increases the generalization's probability in proportion to the inverse of the probability that this new raven would have been black. In other words, in as much as, without presupposing the general rule that all ravens are black, the positive result of the experiment is not guaranteed in advance – this result confirms the rule. By the same token, any additional non-black object listed by the second investigator increases the hypothesis' probability in as much as we have no guarantee in advance of its not being a raven. But here we practically do have such a guarantee, for we can be sure of not encountering ravens by remarking objects in our rooms. This is due to our background knowledge of where ravens are not likely to be. Though, logically, an unexpected encounter, say in one's closet, is possible, its probability is so small as to be negligible. For all practical purposes, the lazy investigator's results do not increase the hypothesis' probability.

Our different rating of the two methods is thus traced back to our background knowledge. One can go one step further and show that without this knowledge there is no general reason for preferring the method of testing ravens for blackness to the method of testing non-blacks for non-ravenness. The relative rating of the two is determined by the following question: without presupposing any rule concerning the colours of ravens, which method has the higher initial probability of discovering a non-black raven? The method which does this is to be preferred. In order to control the effect of background knowledge imagine a well-defined collection of objects containing known percentages of ravens and of non-black objects. Two experimental procedures are available. One – to choose at random until a raven is found whose colour is then tested and recorded; this is repeated, say, n times. (Objects that turn out not to be ravens are ignored and do not enter into the experimental results.) The second – to choose at random until hitting upon a non-black object which is then tested for being a raven; this, again, is repeated n times. A positive result is, in the first experiment, that only 'black' has been recorded; in the second, that only 'non-raven' has been recorded. Which better supports the hypothesis? All depends on the known percentages. If that of ravens is smaller than that of non-blacks then it is the first, but if non-blacks have smaller percentage it is the second. The situation is, mathematically speaking, symmetric, and the larger the difference of percentage, the larger the difference in support. I shall return to a variant of this experiment in §5.

I owe to Simon Kochen the following neat example. Consider the generalization 'Anglo-Saxons are insusceptible to Tay-Sach's disease'.[2] This can be confirmed by checking known cases of the disease. None of them is Anglo-Saxon. But testing Anglo-Saxons and finding no case of the disease constitutes practically no confirmation. For the percentage of diseased people is so small that a random sampling is unlikely to turn up one, even among Jews, where the disease is known to occur. Here is an actual case where our generalization is confirmed by repeatedly observing non-B's which are non-A's; whereas observations of A's that are B's have no value.

The standard solution dispenses with Nicod's rule altogether. In confirming a hypothesis, the significance of a verified consequence is determined by its initial probability. (Later we shall see also the role that the initial probability of the hypothesis plays in this game.) That the consequences happens to be a particular instance of the general statement which constitutes our hypothesis is of no additional import. Apparently, Nicod's rule may yield either true or incorrect results, depending on the case, and we would do better to ignore it and stick to our probabilistic evaluation. Indeed, once the probabilities of the relevant statements are given, all the rest is determined; qualitative confirmation theory and rules such as Nicod's become, at this stage of the game a hinderance.

"Nearly All" Solutions

That nearly all A's are B's need not imply that nearly all B's are A's. By "nearly all" is meant an extremely high proportion (i.e., that the ratio of A's that are B's to A's is at least $1 - \varepsilon$, where ε is very small). The two statements 'all ravens are black' and 'all non-blacks are not ravens' cease to be equivalent, once 'all' is interpreted as 'nearly all'. In which case the puzzle is solved – because our two investigators are not confirming the same hypothesis. Nicod's rule can still be maintained by each. Whereas the first confirms that nearly all ravens are black the second confirms that nearly all non-blacks are not ravens. We reject the second procedure as a confirmation of the first claim for, in fact, it confirms something different (and uninteresting at that). I term a "nearly-all" solution a solution claiming that 'all' in the context of the puzzle should not be interpreted in the strict logical sense but as something weaker.

There are several arguments for a weaker interpretation of 'all'. First, inductive reasoning is an expression of a basic tendency; we expect past regularities to repeat in the future. The tendency is already at work on the level of animal reactions. Conditioned reflexes, expectations formed by habit, association of ideas and inductive generalizing are linked. The expected regularities are expressed in the form of universal generalizations. By asserting them, people often mean that one may safely accept them as guiding rules. It does not entail a strict interpretation of 'all'. One could argue that customary confirmation procedures are applicable where universal laws are understood in this loose sense, but that to apply them to strictly universal statements is a doubtful extrapolation. Arguments in this spirit have been proposed by various thinkers. Carnap held, during the fifties, that universal statements that imply an unbounded number of particular instances (of empirical nature) should get probability zero – a view proposed earlier by Keynes. Consequently such statements cannot be confirmed by any finite piece of evidence. What one usually confirms, Carnap argued, is the reliability of some general law, that is to say – the statement that the law will hold for some large, but finite, number of cases in the foreseeable future. If for 'all A's are B's' one substitutes 'all n A's to be next encountered are B's' then the equivalence needed for the puzzle fails and we get a solution of the "nearly-all" variety.[3]

Customary statistical practice seems to suggest a "nearly-all" interpretation, as well. By standard statistical procedure one could derive, given a sample of ravens, estimates of the relative frequency of black ravens among ravens. One could arrive, say, at $1 - \varepsilon$ as a safe lower bound, where ε is very small. Still, as long as ε is not zero, we are confirming a "nearly-all" statement.

Now, it is quite true that, originally, inductive confirmation is a procedure whereby practical guiding rules get established and that accepting such rules differs from beliefs in universal statements (involving a potential infinity of particular instances). But this, in itself, should not bar the application of the same procedure to strictly interpreted generalizations. An extrapolation it is, but a most natural one. It ought not to be rejected, unless for good reasons. A close look at the arguments against it reveals them to be unfounded. For one thing, Carnap's position concerning universal statements had been more or less forced on him by the fact that his proposed probability functions assigned these statements the value zero. The assignment was not planned ahead; it is a by-product of the very elementary mathematical construction by which Carnap sought to arrive at a probability measure. Yet, probabilities that assign non-zero values to universal statements are easily available (or, for that matter, probabilities having positive values for all consistent sentences belonging to a given countable collection). The thing is a mathematical technicality. Once he realized it, Carnap no longer adhered to his above-mentioned position.

By using probabilities which assign an initial positive value to our general hypothesis, one can arrive not only at $1 - \varepsilon$ as a safe lower bound, but at a high probabilistic value for the strict generalization. The argument from customary statistical practices is beside the point, for these methods are not intended to treat of sharply delineated universal hypotheses. One applies them in situations where answers are sought in terms of frequency estimates and where, for practical purposes, an estimate of $1 - \varepsilon$ is as good as a strict universal assertion, provided that ε is sufficiently small. Strictly interpreted generalizations can, however, be treated within a Bayesian framework and they can be confirmed, provided that their initial probability is positive. The argument that purports to show, by means of rational bets, that generalizations (with an infinite number of empirical instances) have probability zero is erroneous. It is based on a confusion of truth with probability.[4] But I need not go into this here, for Hempel's puzzle can be reconstructed with a finite domain of discourse. Suppose that we consider only physical objects to be observed by human beings within the next fifty years. The generalization becomes 'all ravens to be observed within the next fifty years are black' and the puzzle arises in the same way (see Note 3). One should give good reasons why, in this finitistic context, 'all' cannot be interpreted literally and I do not see how this can be done.

Now, there is a point to replacing 'all' by 'nearly all' but the argument must proceed via the standard solution. To see the connection suppose that ravens and non-black objects have relative frequencies δ_1 and δ_2, respectively, within the total population in our domain of discourse. (For the sake of simplicity assume some very large but finite domain.) Let δ be the relative frequency of non-black ravens. Then the ratio of black ravens to ravens is $1 - \delta/\delta_1$; of non-black non-ravens to non-black objects: $1 - \delta/\delta_2$. That nearly all ravens are black means that δ/δ_1 is very small, that nearly all non-blacks are not ravens – that δ/δ_2 is very small. In as much as δ_1 is smaller than δ_2 the first statement implies the second but not vice versa. The two come closer to the extent that δ_1 and δ_2 approach each other, more precisely – to the extent that δ_1/δ_2 approaches 1. Thus the "nearly-all" solution amounts to asserting that the lazy investigator's method for confirming that all ravens are black is valueless in as much as δ_1 is smaller than δ_2. This, in fact, has been the conclusion of the standard argument. But in the "nearly-all" solution the conclusion is unjustified, whereas in the standard one it is derived from probability theory without having to replace "all" by "nearly all". [A similar analysis applies to the "nearly-all" variant where 'all' is replaced by 'the next *n* so and so's'. Here the two statements are not equivalent, even when $\delta_1 = \delta_2$, but they become sufficiently correlated so that confirmation of one is also confirmation of the other.]

The replacement of 'all' by 'nearly all' yields a useful insight into confirmation procedures. Using the notation just introduced, to say that all ravens are black is the same as saying (assuming some very large but finite population) that the relative frequency of non-black ravens among ravens is 0, i.e., that $\delta/\delta_1 = 0$. Similarly, saying that all non-blacks are not ravens can be put in the form: $\delta/\delta_2 = 0$. The equivalence of the two becomes: $\delta/\delta_1 = 0 \Leftrightarrow \delta/\delta_2 = 0$. The state where $\delta/\delta_i = 0$ can be viewed as a limit state obtaining by letting δ/δ_i approach 0. In the limit, the two states are equivalent. But if we replace '$\delta/\delta_i = 0$' by 'δ/δ_i is very small' the equivalence is no longer valid – it fails to the extent that δ_1/δ_2 is far from 1. Thus, the equivalence of limit states can be *unstable*: it need not necessarily hold throughout a sufficiently large neighbourhood of 0. The moral is that, when confirming by instances, in order that positive instances of one generalization have comparable value as positive instances of a logically equivalent generalization, the equivalence should be *statistically stable.* It should continue to hold where 'all' is replaced by 'at least a ratio of $1 - \varepsilon$ of', with ε varying within a sufficiently large neighbourhood of 0. This is a sound criterion derivable by the standard probabilistic approach.

Natural Predicates Solutions

The recognition of a particular case as an instance of a general law involves already a conceptual organization. I do not mean the Kantian organization whereby raw material is construed as meaningful data; but simply that out of the infinitely many generalizations producible by our logical apparatus we pick as relevant a very limited number. A brown snake is also a non-black object which is not a raven, or a non-white creature which is not a swan or an A which is a B, where 'A' stands for 'a male vertebrate or a female which is not a mammal' and 'B' stands for 'brown or indigo or oviparous'. After ruling out generalizations with known counterexamples we are left with infinitely many, of which very few appear as worth considering. Goodman's puzzle (to be discussed in §3) indicates that our preferences cannot be accounted for by logical structure and by the facts that we know. Nor, I shall argue, can they be derived from unaided formal rules of subjective probability theory. We expect regularities of a very special nature and we tend to classify our objects along very particular lines. Certain categories, or kinds, or, to remain on the linguistic level, certain predicates, are natural; others are not. One can construct ad-hoc generalizations that are formally confirmed by the known evidence but are never taken seriously as candidates for confirmation. When Bacon and Newton spoke of induction as a method of discovering true laws they presupposed the analogy in question to be a natural one. Not every formally construed inductive argument is what Bacon termed "good induction".

Evidently naturalness need not be preserved by logical manipulations, even of the most elementary kind. 'Male', 'female', 'human,' 'reptile' are natural; 'human female or reptile male' is not. Naturalness need not be preserved by negation: 'raven' is natural; 'an object that is not a raven' is not.

If Nicod's rule is to conform with good induction, then it should be modified – an A which is a B confirms the law that all A's are B's provided that *A* and *B* enjoy the non-logical status of naturalness. Since 'non-raven' and 'non-black' do not – the lazy investigator cannot invoke it. I term a solution of this sort a natural-predicates solution. One should perhaps speak of natural pairs of predicates (the pair '*A*', '*B*'), for it is conceivable that naturalness pertains to the combination rather than to each separately, but I shall not go into this. In any case, we are not dealing here with one precise solution, but with a general line. The fact is that the proposal has not got much beyond the stage of a basic idea, though attempts at specific developments have been made.[5] Naturalness has to do with the predicate's place in the totality of our conceptual organization. When we try to spell it out we find ourselves in very deep water. Quine pointed out that our most elementary learning methods are based on innate standards of similarity – we classify objects according to, say, colour or shape, implying that 'round' or 'yellow' are natural. He also pointed out that, in as much as our actions depend on these standards, our innate structuring has consequences for survival; had our classifications been out of tune with nature we would not have survived.[6] The advancement of science changes the organization of our properties-space; more fundamental classifications take the place of the original ones, new natural predicates make their appearance. True as these observations are, they amount to an outline of the phenomenon rather than to an analysis of it. Quine has noted the general connection between the notion of natural kinds and the concepts of dispositional properties, counterfactuals and causal relationship between events. What, I think, has not been generally noted is that natural predicates do not belong exclusively to the empirical domain but are fundamental in purely mathematical frameworks as well. Mathematical knowledge requires a conceptual organization along very particular lines. Here naturalness is connected with equally vague notions such as simplicity and generality. A striking example of a search for natural predicates is provided by the project of defining semantics for computer programs; the very way of organizing material of a purely combinatorial and finite character constitutes here the main problem. In principle, naturalness of predicates is to be traced to the basic organizational patterns of the human brain. We are faced here with a formidable tangle of problems. No wonder that so little has been done.

Ignorant as we are as to what makes a predicate natural, we have a good idea concerning the consequences of this status. Natural predicates are those that figure most in describing the data of a certain domain and in formulating its laws – that is to say, those generalizations believed to be true that occupy key positions in the deduction of other true statements. A universal statement of the form 'all *A*'s are *B*'s', where '*A*' and '*B*' are natural, commands our attention – for we expect true regularities to be thus exhibited. In probabilistic terms this means that, *a priori*, we assign higher probability values to natural hypotheses; these are our preferable candidates for being true.

In some articles 'confirmation' is used only to denote confirmation (by instances) of generalizations based on a natural pair of predicates. Let me emphasize that my usage here is different. In this paper 'confirmation' indicates an appreciable increase in probability that constitutes a step towards accepting the hypothesis as true. (Of this, more later on in §3.).To be sure, in my usage, not all confirmations correspond to inductive generalizations. For an increase in probability may be caused by factors not representing inductive reasoning. Even an increase that results from a particular instance of a general statement need not always be conceived as induction. Existing theory and background knowledge may enter into it to such an extent that it no longer resembles an inductive paradigm.

Modality Arguments

'All ravens are black' can be interpreted as an assertion of necessity – that ravens *must* be black (in this or that sense of "must"). Applying the same policy to the counterpositive 'all non-black objects are not ravens' we get an equivalent assertion. Logical equivalence preserves modalities; if it is necessary that ravens are black then it is necessary that non-blacks are not ravens and vice versa. A modal interpretation does not break the equivalence underlying our puzzle. But it may help to bring out an asymmetry:

Saying that the very nature of being a raven implies blackness is not equivalent to saying that the very nature of not being black implies non-ravenness. Ravens are characterized by a positive common denominator – non-blacks lack such a characterization; they are merely the objects not falling within a certain class. Clearly, we are distinguishing here between natural and unnatural predicates. A solution that is based on the asymmetry revealed in the modal version is, in fact, a natural-predicates solution. This goes also for the causal interpretation of 'all ravens are black': being a raven causes its being black. The asymmetry is obvious.

Being a raven may cause blackness but evidently, non-blackness is not the cause of not being a raven, it is only a symptom. Here, as well, we are within the natural predicates framework. Properties that can play the role of causes are natural. The ruling out of a causal arrow from non-blackness to not being a raven is rooted in the epistemic status of 'non-black, and 'not a raven'.

Modal versions, in particular, assertions of causality, serve also as possible explanations of general phenomena. Here, we are not concerned with the explanatory aspect per se. But the establishing of a general rule and the explaining of it are connected. When possible explanations are available the *a priori* probability of the proposed rule is higher.

There is another, altogether different, modality argument, purporting to dismiss the puzzle altogether. The argument (advanced by Lakatos and others) consists of two claims. First, that, as a scientific hypothesis, 'all ravens are black' ought to be interpreted as asserting a necessity, since science aims at necessary laws, not at mere general facts. Second, that looking for ravens in order to observe their colour is of no value for establishing the modal assertion; a much more radical experiment is needed. So, our hard-working investigator fares no better than the lazy one. Proponents of their view reject confirmation by instances altogether as well as the application of probability to situations involving scientific hypotheses. This view is erroneous. Let h be our generalization and let h^* be its modal counterpart. Evidently h^* logically implies h but not vice versa. If h were refuted by a counterexample, h^* would be refuted as well. For this reason alone observing ravens is relevant to accepting h^*; it is an experiment, whose results are not guaranteed in advance, that may refute h^*. To be sure, h does not logically imply h^*. Yet the truth of h would grant h^* considerable plausibility and, under certain circumstances, would make it highly probable. Knowing that millions of ravens dispersed all over the world are invariably black, would we not conclude that this is not a mere accident? The bearing out of a general statement by all known instances, where these constitute a large enough class, is taken as a mark of lawfulness, i.e., of necessity. This is best exemplified when no elaborate scientific theory intervenes. Mortality has been regarded as essential to human nature by virtue of the unrefuted generalization that all humans are mortal. (Even now, no theoretical derivation of the law exists, though in recent years speculations have been made connecting it with the limited divisibility of cells.) The situation is more complex when various explanatory models are provided by existing theory. For one thing, the very concept of necessity becomes differentiated and should be clarified. Anyone proposing h^* should specify what sort of necessity he intends. Does h^* claim that non-black ravens are impossible even by means of future genetic engineering? Or does it only take into account manipulations that preserve the genetic set-up? Would h^* be true if it turned out that blackness is for ravens an essential survival feature? The best way of confirming h^* would depend on its exact meaning and on existing theory. That confirmation by instances may not be the best way has no bearing on the puzzle. It suffices that, in general, it does have confirmatory value towards establishing scientifically significant results.

Failure to appreciate the commonplace distinction between probability as a measure of frequency and probability as a degree of belief may have created the wrong impression that probability (as degree of belief) applies only to non-modal statements. Actually, any statement having a truth-value is a legitimate candidate for probabilistic assessment. One can meaningfully speak of his probability that a causal connection exists between two events, or that certain facts are necessary, provided that, in the given context, 'causal connection', or 'necessary' are sufficiently clear. General probability laws, such as Bayes', apply irrespective of the nature of the statements in question. Hence, we can consider the confirmatory value that the evidence, e (gathered by our hard-working investigator) has toward establishing some modal version, h^*, of h. This value to be defined in §3 (under the name "confirmation rate") depends not only on the initial probability of e, but also on that of our hypothesis. The value of e for h^* can be significant, though, in general, it will be smaller than its value for h.

3. SUBJECTIVE PROBABILITY

The framework of subjective probability provides for a systematic treatment of situations where knowledge is uncertain. It is based on a numerical representation of degrees of belief (or credibilities, or plausibilities). This does not necessarily imply computations of particular real numbers, as carried out in concrete applications; but a real-valued function representing a system of beliefs is presupposed, and the relationship between various hypotheses and the known evidence is analysed in terms of this function. 'Subjective' in this context is a technical term indicating probability as a measure of degree of belief, as distinguished from objective frequencies existing in nature. (No doubt, knowledge of the objective frequencies will often determine one's degree of belief; one's probability that a particular unseen card, that has been drawn at random, is red, is also his estimate of the ratio of red cards in the pack.) Like other philosophical enterprises the probability framework begins by analysing actual practice and ends by establishing normative criteria. The basic rules, to be presently stated, are well known. Let me remark first that the mathematical parlance according to which probabilities are assigned to "events" is merely a manner of speaking in which the term 'event' does not have its ordinary, or its philosophical, significations. In fact, any state of affairs can be an "event", no matter whether it is realized or not, including the so-called "empty event" – a state of affairs described by a contradictory statement. So probabilities are assigned, roughly speaking, to what is described by factual statements. In order to avoid undue philosophical complications let us construe the probability function as an assignment of values to *sentences*. The value measures the degree of belief in the sentence's truth and the probability function is supposed to reflect a coherent system of beliefs. The class of sentences over which the function is defined should include sentences stating whatever hypotheses and describing whatever pieces of evidence we want to analyse. It should be closed under the usual sentential connectives (negation $\neg$, disjunction $\vee$, conjunction $\wedge$, etc.). The probability function has values in the closed interval [0, 1] and the basic requirements are:

(1) Logically equivalent sentences are assigned the same probability.
(2) Tautologies are assigned probability 1.
(3) The probability of a disjunction of two mutually exclusive sentences is the sum of the probabilities of the sentences.

Let $p(a)$ be the probability of the sentence a. Then the three rules can be formally written as:

(1) If $\vdash a \leftrightarrow b$ then $p(a) = p(b)$
(2) If $\vdash a$ then $p(a) = 1$
(3) If $\vdash \neg(a \wedge b)$ then $p(a \vee b) = p(a) + p(b)$.

Other elementary well-known rules are consequences of these. The rules are constitutive of the framework. They determine the meaning of the probability function in the same sense that modus ponens and the other rules of classical logic determine the meaning of the connectives. One can "justify" them by interpreting one's subjective probabilities in terms of the betting odds that he is prepared to accept; or by showing that, in the case where the subjective probabilities are identified with known (or estimated) objective frequencies, the rules are derivable. In fact, all the justifications are arguments purporting to show that this is the right sort of system for treating of degree of belief. Just as one might give various arguments showing that Peano's axioms are the right sort for treating of the intuitive concept of natural number.

One's probabilities usually change upon his acquiring new factual information. If $p(a)$ is the probability one accords to a, then $p(a/e)$ is defined as the probability that he should accord to a in a state of knowledge obtained by adding e to the set of sentences known as true; $p(a/e)$ is said to be the *conditional probability of a, given e*. A fourth basic rule determines (in the case that $p(e) \neq 0$) the conditional probability in terms of the original values of the function:

(4) The conditional probability of one sentence, given another, is the ratio of the probability of the conjunction of the two to the probability of the second. In symbols: $p(a/e) = \dfrac{p(a \wedge e)}{p(e)}$.

This rule can be argued for much in the same way as the preceding ones. (Note that, had we regarded (4) as a definition of $p(a/e)$, we would need a rule stating that $p(a/e)$ is indeed the probability one should accord to a upon the adding of e to the known true sentences.) The remarkable consequence of (4) is that a probability function, defined over a class of sentences, determines already any probability function that will arise upon the acquiring of new factual information (provided that the information is describable by a sentence in the class which, initially, is accorded a non-zero probability).

An immediate consequence of (4) is Bayes' law:

(5) $$p(h/e) = p(h)\frac{p(e/h)}{p(e)}.$$

The ratio $p(e/h)/p(e)$ is the factor by which the initial probability of the hypothesis, h, changes as a result of the evidence. If e is implied by h we have $p(e/h) = 1$ and the factor becomes $1/p(e)$. Note, however, that the ratio of probabilities is not necessarily a good measure of the confirmatory value of e. For often we mean by 'confirmation' an increase in our readiness to accept the hypothesis (in one of the senses of acceptance). What counts is the extent to which the evidence causes the probability of h to come nearer to 1; an increase from 0.6 to 0.9 is highly significant, an increase from 10^{-4} to 10^{-3} is usually not, although the ratio of the first is 1.5, compared to a ratio 10 of the second. We can try to take the difference of probabilities

$$p(h/e) - p(h)$$

as an indicator. A better one is the ratio of distances from 1, what I call the *confirmation rate*, or $\mathrm{Cr}(h, e)$,

$$\mathrm{Cr}(h, e) =_{\mathrm{DF}} \frac{1 - p(h)}{1 - p(h/e)}.$$

Other candidates are conceivable, but each is to be applied with discretion; for example, $\mathrm{Cr}(h, e)$ is not a good indicator if $p(h)$ is already very near 1, so that further increase is of no great consequence. Precautions of this sort are inevitable, for the significance of an increase in subjective probability is very much context dependent.

The standard solution of §2 can be now formally described. We let '$e_1(n)$' stand for 'the first n ravens observed by the first investigator were black', and '$e_2(n)$' for 'the first n non-black objects observed by the second investigator were not ravens'. For reasons given in §2, $p(e_2(n))$ is practically 1; hence, for all practical purposes, $p(h/e_2(n))$ is indistinguishable from $p(h)$, where 'h' stands for 'all ravens are black'. By comparing the values of $\mathrm{Cr}(h, e_i(n))$ for $i = 1, 2$ we can arrive at the relative rating of the two pieces of evidence. As remarked in §2, the standard solution bypasses Nicod's rule. It bypasses also the issue of natural predicates. The naturalness aspect is irrelevant when a probability function, or a basic outline of one, has been accepted. But it is highly relevant at the stage of setting the function up. In connection with this, I shall now take a short detour and look into some problematic aspects of the subjective probability framework.

A probability that is defined over an extensive class of sentences is, by nature, a mathematical idealization. 'Subjective' does not indicate, in this context, a thorough acquaintance with the probability function – even when it is said to represent one's own system of beliefs. For one may err, or one may be unaware of the implications of his basic assumptions. Just as in the deductive domain one may believe things which, in fact, are inconsistent, one may go completely wrong in estimates of probabilities. There has been considerable research on the fallacies of everyday probabilistic thinking. So a subjective probability-function is an idealization which reflects, as best possible, a basic pattern of beliefs while avoiding inconsistencies. In order to get an idea on one's function, he should analyse his intuitions and, in case of conflict, try to find the clearer or more fundamental ones. Techniques have been developed, utilizing betting situations, in order to bring out more clearly the details of one's assessments. Other Gedanken experiments can be utilized as well.[7] Yet, when all is done, quite a few details are left obscure. The function is a theoretical posit whose mathematical content overextends its concrete interpretation. The difficulty of pinpointing the complete subjective probability-distribution has been an argument of non-Bayesian statisticians against this very concept. The argument is unjustified, for not all details of theoretical constructs should have concrete correlates in the situations to which they apply. Bayesian statistics has been doing very well, for, in concrete applications, one need not know all the details of the function. Certain estimates will suffice, concerning which we may have clearcut intuitions. Furthermore, those features which determine our judgements in a given situation may be shared by different functions which represent different beliefs. Thus, a consensus is arrived at though the underlying beliefs vary.[8]

In bringing together different systems of beliefs the commonly shared accumulated evidence plays a crucial role. A probability $p(\ )$ can be regarded as a conditional probability $p_0(\ /e)$, where $p_0(\ )$ is a function at some earlier state of knowledge and e is the evidence accumulated since then. In such a context p_0 is said to be the *prior probability*, or, for short, the *prior*. Evidently, this concept is relative; any function can be regarded as a prior with respect to functions arising out of it upon the accumulating of further evidence. Still, one can imagine a fixed original prior (or family of such[8]) and represent our actual function (or family of functions) as the resulting conditional probability, given all the evidence accumulated by humans. This is not to claim that the hypothetical prior corresponds to some real historical state. The point of introducing it is that it determines, through (4), future possible developments. It sums up neatly our ways of reacting to new evidence. To represent our present probabilities as being the conditional probabilities of some original prior, relative to a very large amount of collected evidence, is to represent our beliefs as a combined outcome of our basic ways of learning from experience and of the objective facts that we have come across. Since the facts are forced on us, whatever "justification" our belief-system needs, if any, is a "justification" of the prior.

One can add to the basic rules given above, additional, quite reasonable, requirements which will restrict the family of possible priors.[9] There are also mathematical theorems showing that within certain families any two priors will give rise to converging conditional probabilities as the evidence grows; that is to say, although $p(\ )$ and $p'(\ )$ may be very different, $p(\ /e)$ and $p'(\ /e)$ will come in the long run, as e represents larger and larger chunks of data, arbitrarily close. If this convergence is uniform with respect to all pairs of functions in our family, then, given sufficient data, every two priors will lead approximately to the same probability estimates. This situation often takes places in statistical applications of subjective probability. But then the families of priors which the statistician presupposes reflect our common-sense and are considerably narrower than families defined by formal conditions.

By a "formal condition" I mean one that links the behaviour of the function with the logical relations of the sentences over which it is defined. ((1), (2), (3), as well as the additional requirements just mentioned (cf. Note 9) are formal conditions.) One can achieve mutual convergence of any pair of functions by imposing certain formal conditions, but one cannot achieve a uniform rate of convergence for all pairs in the family. For, given any general hypothesis that has been so far borne out, one can construct an ad-hoc rival that conforms with the known evidence, but disagrees with the first as to the future. One prior can favour the first hypothesis against the second, another – the second against the first. The probabilities assigned by such a pair of priors to these two hypotheses differ in the extreme; and so do the respective conditional probabilities as long as the evidence is compatible with both. The mutual convergence may start only at a stage where evidence does not conform with the two hypotheses. The point is neatly exemplified in Goodman's puzzle, where the standard law that all emeralds are at all times green is confronted with a non-standard rival that all emeralds are at all times grue. (Grue at time t is defined as being green if $t <$ A.D. 2000, blue if $t \geqslant$ A.D. 2000.) Since up to A.D. 2000 'green' and 'grue' have the same extension the positive instances of the two hypotheses coincide up to that date. Let h_1 be the standard hypothesis, h_2 the non-standard rival and e the evidence that all observed emeralds have been until now green (thus, also grue). We accept, given e, h_1 and no-one in his right mind will assign h_2 an appreciable chance; hence, $p(h_1/e)$ is near to 1 and $p(h_2/e)$ is close to 0. Since h_i implies e, for both $i = 1, 2$, $p(h_i/e) = [1/p(e)]p(h_i)$. Hence $p(h_2/e)/p(h_1/e) = p(h_2)/p(h_1)$, meaning that the prior probability of h_2 is insignificant compared with that of h_1. The conditional probabilities, given e, are obtained through multiplying each by $1/p(e)$; this brings the probability of h_1 near 1, while that of h_2 remains insignificant.[10] Now, one can conceive another prior, p', which reverses the roles of h_1 and h_2, i.e., $p'(h_1)/p'(h_2) = p(h_2)/p(h_1)$. Starting with p' we shall, given e, accept h_2 and reject h_1. The two conditional probabilities $p(h_1/e)$ and $p'(h_1/e)$ (or $p(h_2/e)$ and $p'(h_2/e)$) will start to converge towards each other (where e consists of reported observations of emeralds) only if e contains reports pertaining to times after A.D. 2000.

Non-standard priors such as p' must be excluded if our family of priors is to represent faithfully our ways of learning from experience. In fact, no statistician will even consider them. But this exclusion cannot be effected through formal-structural conditions. For in point of logical structure h_1 and h_2 play symmetric roles. One cannot set up formal requirements that will force us to favour h_1 over h_2. We could try to construct the probability function from the bottom upwards – starting from atomic sentences, their negations and conjunctions of such and proceeding to more and more complicated sentences. We could require that our function conform with certain intuitions concerning inductive reasoning; for example, that the conditional probability of $P(b)$ (where 'P' is some primitive predicate) given a conjunction of the form $P(a_1) \wedge \cdots \wedge P(a_n)$, increases fast enough with n. Under these circumstances the choice of primitive predicates (as well as the descriptions, or labelings, by which individual objects are identified) becomes crucial. Choosing 'green' and 'blue' as primitives and defining 'grue' in terms of these would lead to a probabilistic preference of h_1. But, as Goodman observed, there is no logical-structural reason for not defining 'green' and 'blue' in terms of 'grue' and 'bleen'.

A systematic account of how we come to use a prior of a certain kind must resort to epistemic analysis. At this juncture the concept of a natural predicate becomes important. For the prior expresses our innate expectations of regularities – those generalizations that we tend to make *a priori*, the "good" inductions.

The status of a natural predicate may and often does change with accumulated evidence. Given certain evidence, the conditional probability may indicate expected regularities that are most clearly expressible in terms of originally complex constructions. Se we take these constructions and treat them as primitives, a step that is usually signified by the introduction of a single primitive term for what used to be a lengthy description. The primitive basis gets changed. But our original basis, or, what comes to the same, our original prior, cannot be accounted for in this way.

4. SOME OBSERVATIONS CONCERNING NATURAL PREDICATES AND FREQUENCIES

There is sometimes a tendency to correlate low frequency with naturalness. This, I believe, is rooted in the fact that at the rudimentary stage of language-formation, predicates correspond to directly perceived properties which, in order to be perceived, should appear on a wider contrasting background. One becomes aware of a certain sound when it interrupts a longer period of silence, or of a different noise. Pavlov trained his dogs to associate food with the sound of ringing. But had he used a dog that had been subjected since birth to an almost continuous ringing, Pavlov should have used silences as stimuli. In the language of Pavlov's dogs, 'ringing' would be a natural predicate, but in the language of this last dog 'no-ringing' would take its place. The correlation ceases however to obtain on the level of elementary concept formation. Before arguing the point let me observe that even if low frequency were correlated with naturalness this would not introduce the concept of natural predicates into the probabilistic solution. For we estimate ravens as less likely to be encountered than non-black objects by virtue of past experiences, not by virtue of 'raven''s naturalness. (By contrast, the assignment of probabilities in the case of Goodman's puzzle *is* explainable by epistemic factors.)

Natural similarities can be grasped even when the instances in question make up almost all of the perceived space. Silence is identified by contrast with noise, but once recognized, the similarity of all silent moments is evident. A silence that is but a part of a long soundless stretch will go unnoticed, but a natural predicate, 'silence', is available to characterize it. Imagine somebody undergoing his first visual experience in a room where 85% of all coloured area (or all coloured objects, or whatever measure you prefer) is uniform white, with green indigo and purple making up the other 15%. Will he not classify all white patches in one group because of their recognizable common denominator? 'Not-white', though true of only 15% of his surroundings, cannot constitute in his organization a natural predicate, because such a common denominator is lacking. 'Male' and 'female' are natural predicates whose respective extensions comprise, each, about 50% of the domain of discourse (this being the domain of humans, or of mammals, or any similar domain). 'Neither male nor female' indicates a minute class. Yet it does not possess the same degree of naturalness, as it lacks the sort of common quality that characterizes members of the other two. On a higher level of abstraction we have 'physical object', 'time interval', 'spatio-temporal location' with which low frequency can in no way be associated. In science we have 'atom', 'electron' and their like and in primitive science we have 'air' whose all-pervading presence was in no way an obstacle to its being thought of as a natural kind – a single fundamental substance. One concludes that in a nightmare world where ravens are as frequent as non-black objects are now and where non-black objects are as rare as our ravens, 'raven' would still be natural, 'non-black' – not.

Another observation, to be used in §5, concerning the irrelevance of relative size, has to do with the distribution of a property, say B, within the class of all A's; i.e., the frequency of B within A. The actual distribution being unknown, we have probabilities assigned to the possible frequencies, e.g., the probability that the relative frequency of B within A is bigger than 0.5. (Or, if one does not wish to commit oneself to an idealized prior, we have certain initial ideas of what the frequency of B is likely to be.) The principle of irrelevance of size is that knowledge of the frequency of A should not affect these probabilities. For example, knowing how rare or how frequent ravens are is irrelevant for estimating the percentage of black ravens within the raven population.[11] This principle, plausible as it appears, does not always obtain. It is conceivable that, by virtue of existing theory, a correlation is known between the relative sizes of classes in a certain family and the percentage of B's within each; say, the more numerous a species the higher the percentage of its blue-eyed members. But, not having such a background theory, given only a natural classification into A's and non-A's and into B's and non-B's, there is no way in which knowing A's frequency can be relevant to estimating B's frequency within A. Note that if the classification is not along natural lines, such as raven–non-raven, black–non-black, the principle may fail. Definitions of A and B can be set up so as to make A's frequency informative with respect to B's frequency in A. (The reader may try this as an exercise.) On the other hand, the principle depends not so much on the predicate's naturalness as on the naturalness of the division lines. It should hold also for non-A and non-B. Knowing the relative size of the class of non-ravens does not affect our estimates concerning the distribution of blackness, or for that matter of non-blackness, among them. The principle can be generalized as follows: Let D be a subclass of the class of all non-A's defined using a natural division, then the size-ratio (or frequency ratio) of A to D is irrelevant for estimates concerning the relative frequency of B within A. For example, knowing the size-ratio of the class of ravens to the class of all objects that are neither raven nor black is irrelevant for probabilistic estimates of the frequency of blackness among ravens.

This last statement will be used in §5. (We need it in order to facilitate the mathematics. The final conclusions would have been the same in any case.) Note that here, as well, what matters is the naturalness of the division-lines. The statement would still hold had we replaced everywhere 'raven' and 'black' by their negations.

5. FINAL ANALYSIS

Even without carrying out a quantitative analysis one should not expect Hempel's puzzle to be solved by invoking the concept of natural predicates. It and Goodman's puzzle do not belong together. Let us compare them by casting each in the form of a question. Letting e be the evidence that, hitherto, all observed emeralds have been green and letting h_1 and h_2 be, respectively, the standard and non-standard hypotheses, Goodman's query becomes:

Why does e, which consists of positive instances of h_2, as well as of h_1, make us accept h_1 and not h_2?

Similarly, (if e_1 and e_2 are, respectively, the evidences collected by the hardworking and the lazy investigators and if h is the hypothesis that all ravens are black) Hempel's query is:

Why does e_1 confirm h, whereas e_2 ,which consists of positive instances of a logically equivalent hypothesis, does not?

The probabilistic answer to the first query is that the initial probability of h_1 is incomparably bigger than that of h_2. This last fact is a feature of our prior distribution and when we try to account for it we are faced with epistemic problems concerning the organization of our properties-space. At this stage some concept of natural predicates enters the picture. Now, the answer to the second query is that e_1's probability, $p(e_1)$, is very much smaller than $p(e_2)$ – a value which is so near 1 as to render $1/p(e_2)$, for all practical purposes, 1. Before proceeding note that this last explanation cannot be bypassed. One *cannot* argue at this juncture that e_2 is of little value simply because it consists of positive instances of a version in which the predicates are not natural. For one will have to show how this affects the value of $p(h/e_2)$ as compared with that of $p(h/e_1)$. If a implies b, $p(a/b)$ equals $[1/p(b)]p(a)$; hence in explaining why the value is high, or low, one must resort either to $p(b)$ or to $p(a)$, or to both. In the case of Goodman's puzzle, $p(h_1/e)$ is compared with $p(h_2/e)$; consequently it comes to comparing $p(h_1)$ with $p(h_2)$. In our present case $p(h/e_1)$ is contrasted with $p(h/e_2)$: one must therefore carry out a comparison of $p(e_1)$ with $p(e_2)$. We may, as in the first case, push our question further: why is it that $p(e_2)$ is so much nearer to 1 than $p(e_1)$? The answer given in §2 is simple and does not derive from epistemic fundamentals.

The naturalness aspect is expressed in our prior probability distribution. It is difficult to see, without a detailed analysis, how this influences $p(e_i)$. But the main contribution of this aspect is due to the fact that, in most cases, not $1/p(e)$ but $\mathrm{Cr}(h, e)$ is a good indicator of the confirmatory value of e. As defined in §3, $\mathrm{Cr}(h, e)$ is $[1 - p(h)]/[1 - p(h/e)]$ and if $p(h)$ is very small compared to $p(e)$ this expression is practically 1, i.e., no appreciable confirmation takes place. This means that h must have appreciable probability to start with in order to be confirmable (by any ordinary kind of evidence). For example, the hypothesis that all ravens are black before A.D. 2000 and white afterwards is confirmed neither by evidence of black pre-2000 ravens nor by non-black pre-2000 non-ravens. To be sure, assuming ravens to be less frequent than non-black objects, the first is the more relevant evidence but the difference on the $\mathrm{Cr}(h, e)$-scale is minute. Thus, naturalness is needed in order to give h some initial plausibility, without which the difference between e_1 and e_2 would not emerge. The higher $p(h)$ the more marked this difference, provided that $p(e_1)$ is different from $p(e_2)$. But which evidence is preferable is determined by frequency estimates.

There is still another, more complex, effect of naturalness which I shall discuss at the end of §5 (see also the postscript). Its relative weight turns out to be minor.

Let us return to the experiment of §2. Specimens are drawn at random from a collection C of various objects. Say that they are drawn by a laboratory assistant (operating some random choice device) and that he can be either instructed to pass on for further testing only those drawn objects that he checked and found to be ravens, or he can be instructed to pass on only the drawn non-black ones. In the first case the ravens are tested for being black, in the second – the non-blacks are tested for being non-ravens. Let $e_1(n)$ be evidence to the effect that all the first n tests of ravens in the first procedure yielded the verdict "black" and let $e_2(n)$ be the analogous evidence with respect to the second procedure (i.e. that the first n tests of non-blacks yielded a "non-raven" – verdict). As before, h is our hypothesis. Assume that our knowledge of C is such that, as far as being black is concerned, our probability is the same for a raven drawn at random from C and a raven drawn randomly in the world; furthermore – that the same is true of a sequence of independent random drawings. This condition is met if C is assumed to contain all ravens.[12] But much less is needed. It suffices that we regard the ravens in C as a random sample with respect to the property of being black; drawing randomly from a random sample is like drawing randomly from the parent population.

Let 'w_{BR}' stand for 'the world ratio of black ravens to ravens.' Given that w_{BR} equals x, the probability that a randomly drawn raven is black is x. The probability that all of the n independent drawings yield black specimen is therefore[13] x^n. Thus $p(e_1(n)/w_{BR} = x) = x^n$.

Let 'D' stand for 'object that is neither a raven nor black' and 'C_{RD}' – for 'the ratio within C of ravens to D's'. If $C_{RD} = \delta$ and if y is the ratio in C of black ravens to ravens, then an easy calculation shows that the ratio in C

of D's to non-black objects is $1/[1 + \delta(1 - y)]$. (If C contains m ravens than $m \cdot y$ of them are black and the rest $m(1 - y)$ are non-black; as there are m/δ D's in C, the total number of non-blacks in C is $m/\delta + m(1 - y)$. Dividing m/δ by the last number yields $1/[1 + \delta(1 - y)]$.) Given that $C_{RD} = \delta$ and that y is the ratio of black ravens to ravens in C, the probability of $e_2(1)$ is $1/[1 + \delta(1 - y)]$. For $e_2(n)$, with n independent drawings of non-black objects we get[14] $(1/[1 + \delta(1 - y)])^n$. If C is known to contain all ravens in the world then y is the actual world ratio of black ravens to ravens; hence, given that $w_{BR} = x$ and that $C_{RD} = \delta$, we get $(1 + \delta(1 - x))^{-n}$ as the probability of $e_2(n)$. In general, x is the expected value of y, though the actual y may not coincide with it. There will be a probability distribution over the possible values of y and $p(e_2(n)/C_{RD} = \delta \wedge w_{BR} = x)$ is obtained by integrating. But it can be shown that we may safely take $(1 + \delta(1 - x))^{-n}$ as our value, (even without assuming that C contains all ravens) for, under our suppositions, the final outcome will be practically the same.[15] All in all we have:

$$p(e_1(n)/w_{BR} = x) = x^n$$
$$p(e_2(n)/C_{RD} = \delta \wedge w_{BR} = x) = (1 + \delta(1 - x))^{-n}.$$

If C is the domain of all physical objects (or any natural domain including the class of all ravens) then, by the principle of irrelevance of size stated in §4, our probability that w_{BR} is, say, between a and b is not changed by the information that the ratio of ravens to non-black non-ravens is δ. I assume that this is also true of the particular C of the experiment. This is an additional natural assumption on C which holds, unless one takes special care to define C in such a way as to make the value of C_{RD} informative for w_{BR}. For instance, if C is the class of objects in a certain geographical location then our assumption holds, by virtue of the same arguments that support the principle of irrelevance. I shall regard the equality $C_{RD} = \delta$ as background knowledge and I shall use 'p_δ' for the probability function, given that knowledge, i.e., $p_\delta(a) = p(a/C_{RD} = \delta)$, implying that $p_\delta(a/b) = p(a/b \wedge C_{RD} = \delta)$. Let Cr_δ be the confirmation rate for the probability p_δ. By our assumptions we have:

$$p_\delta(a \leqslant w_{BR} \leqslant b) = p(a \leqslant w_{BR} \leqslant b),$$
$$p_\delta(e_1(n)/w_{BR} = x) = x^n,$$
$$p_\delta(e_2(n)/w_{BR} = x) = (1 + \delta(1 - x))^{-n}.$$

The second equality holds because the ravens in C are regarded as a random sample, as far as being black is concerned, and the information concerning C_{RD} does not affect this state.

$p_\delta(e_i(n))$ is now obtainable by standard procedure as a weighted sum, where each $p_\delta(e_i(n)/w_{BR} = x)$ is multiplied by a weight representing the probability that $w_{BR} = x$. This can be put in the form of an integral of $p_\delta(e_i(n)/w_{BR} = x)$ with respect to the measure induced on [0, 1] by p_δ, the measure of any set being the p_δ-probability that w_{BR} belong to it. Now p and p_δ induce the same measure, so letting 'p' denote also the measure induced by p we have:

$$p_\delta(e_1(n)) = \int_0^1 x^n \,\mathrm{d}p(x)$$

$$p_\delta(e_2(n)) = \int_0^1 (1 + \delta(1 - x))^{-n} \,\mathrm{d}p(x).$$

For $\delta \leqslant 1$, x^n is, for all x in [0, 1] not bigger than $(1 + \delta(1 - x)^{-n}$; equality takes place only for $x = 1$. It is expected that (for $\delta \leqslant 1$) $p_\delta(e_2(n))$ is bigger than $p_\delta(e_1(n))$, implying that $e_2(n)$, having the higher *a priori* probability, is of smaller value. As δ decreases $(1 + \delta(1 - x))^n$ approaches 1 and so does $p_\delta(e_2(n))$. For very small δ's $p_\delta(e_2(n))$ is practically 1 and $e_2(n)$ is valueless. The details depend, however, on the measure induced by p on [0, 1] and it is here that the naturalness of predicates enter.

Let $\alpha = p(h)$. As h is equivalent to the equality $w_{BR} = 1$, α is the measure of the set consisting of the single point 1, $\alpha = p(\{1\})$. Let $\beta = p(\{0\})$, i.e., β is the *a priori* probability that all ravens are non-black. To regard 'raven' as a natural predicate is to expect ravens to exhibit uniform patterns; in particular, the naturalness of the pair 'raven' – 'black' means an expectation that ravens will behave alike with respect to being black. The stronger this expectation the higher the values of both α and β (for if ravens behave alike they are either all black or all non-black). Hence I can use α and β as indicators of naturalness.[16] In the numerical examples to be worked out I take $\alpha = 0.2$, $\beta = 0.4$ to mark a situation where the pair 'raven' – 'black' is regarded as highly natural, $\alpha = 0.04$, $\beta = 0.1$ to mark low naturalness and $\alpha = 0.1$, $\beta = 0.2$ as an intermediate case. (I put $\beta = 2\alpha$, because, without any information on the colour of ravens and with various colourings as possibilities, it is *a priori* more probable that all are not black than that all are black. The factor 2 has, of course, no importance – the resulting picture would have been the same with any other factor.)

The rest of the probability mass is $1 - (\alpha + \beta)$. It is distributed among the other possible values of w_{BR}. These are $1/\mathcal{N}, 2/\mathcal{N}, \ldots, (\mathcal{N} - 1)/\mathcal{N}$, where $\mathcal{N}$ is the total number of ravens in the world. $\mathcal{N}$ is unknown, but we have a rough idea that it is sufficiently large, say more than 10^4. For many purposes we can replace a discrete distribution over the points $i/\mathcal{N}$ by a continuous one which "smears" the mass over the interval (0, 1). Mathematically, this is more convenient and the final values are practically the same.[17] I should also remark that the conclusions concerning the values of $e_1(n)$ and $e_2(n)$ and the relative importance of α, β and δ would be the same had I used a discrete distribution, even with a small $\mathcal{N}$. In the numerical example I use a uniform distribution over (0, 1), that is to say, the probability that w_{BR} is in (a, b), where $0 \leqslant a < b \leqslant 1$, is proportional to $b - a$. This means that, in our judgement, no frequency in (0, 1) is considered as *a priori* more likely than others. This is not necessarily the only plausible prior. One could, for example, choose a cup-shaped density function that gives higher weight to intervals near the extremes. The argument for it is that the naturalness of 'raven' – 'black' means also that we expect most ravens to share the same colouring, even assuming that some are black and some are not. However, this would lead to essentially the same picture, as far as comparing $e_1(n)$ and $e_2(n)$ is concerned. The numerical values would be different but δ will continue to play the crucial role and our conclusions will not be affected. Certain traits are shared by all priors, or all reasonable ones. For example, if $\delta \leqslant 1$ then $\mathrm{Cr}_\delta(h, e_2(n))$ is always smaller than $\mathrm{Cr}_\delta(h, e_1(n))$. Their ratio decreases with δ. For all priors within a certain wide and natural family[18] it approaches δ as n grows. This means that, for $\delta > 1$, the evidence $e_2(n)$ has, in the long run, higher confirmatory value than $e_1(n)$. The numerical values to be given are a good indicator to the phenomenon in general.

Our prior gives probability α to $w_{BR} = 1$, β to $w_{BR} = 0$ and has a constant density function of height $1 - \alpha - \beta$ over the interval (0, 1). Hence

$$p_\delta(e_1(n)) = \alpha + \int_0^1 (1 - \alpha - \beta)\cdot x^n \,\mathrm{d}x$$

$$= \alpha + \frac{1}{n+1}(1 - \alpha - \beta);$$

$$p_\delta(e_2(n)) = \alpha + \beta\cdot(1+\delta)^{-n} + \int_0^1 (1-\alpha-\beta)$$

$$\times\ (1+\delta(1-x))^{-n}\,dx = \alpha + \beta(1+\delta)^{-n}$$

$$+ \frac{1-\alpha-\beta}{\delta(n-1)}(1-(1+\delta)^{-n+1}), \qquad \text{for } n \geqslant 2;$$

for $n = 1$ the last summand should be replaced by $(1-\alpha-\beta)/\delta \log(1+\delta)$.

The values of $p_\delta(h/e_i(n))$ and of $\mathrm{Cr}_\delta(h, e_i(n))$ are now obtained by substituting the expressions for $p_\delta(e_i(n))$ in the equalities

$$p_\delta(h/e_i(n)) = \frac{p_\delta(h)}{p_\delta(e_i(n))} = \alpha\cdot(p_\delta(e_i(n))^{-1};$$

$$\mathrm{Cr}_\delta(h/e_i(n)) = \frac{1-p_\delta(h)}{1-p_\delta(h/e_i(n))} = \frac{1-\alpha}{1-\alpha\cdot(p_\delta(e_i(n)))^{-1}}.$$

The values for $e_1(n)$ do not depend on δ. They are given, for $n = 1, 2, 4, \ldots, 1024$, in our 3 cases ($\alpha = 0.2\ \beta = 0.4$, $\alpha = 0.1\ \beta = 0.2$, $\alpha = 0.05\ \beta = 0.1$) in Tables 1.1–1.3, where they are confronted with the respective values for $e_2(n)$ for the case $\delta = 0.001$. This δ means that our class has only one raven per thousand non-black non-ravens. (In Hempel's puzzle the ratio should be smaller, for C consists of objects in the lazy investigator's apartment, or neighbourhood.) It is evident that, unless n reaches the range of the hundreds, $e_2(n)$ has practically no confirmatory value. (For smaller δ's even this range of n would not suffice.) Both $e_1(n)$ and $e_2(n)$ gain in confirmatory value when α and β are increased, i.e., when 'raven' – 'black' is taken to be more natural. But $e_1(n)$ gains more and this can be shown by considering the ratio of $\mathrm{Cr}_\delta(h, e_1(n))$ to $\mathrm{Cr}_\delta(h, e_2(n))$. This ratio, roughly speaking, indicates the extent to which $e_1(n)$ is superior to $e_2(n)$. It is higher for bigger α and β. Here lies the effect of naturalness.

To exhibit the frequency effect we take the case $\alpha = 0.1$, $\beta = 0.2$ for various δ's, $\delta = 0.05$, $\delta = 0.1$, $\delta = 0.5$ and $\delta = 1$. As δ increases $e_2(n)$ gains rapidly in confirmatory value and the gap between $e_1(n)$ and $e_2(n)$ decreases. For $\delta = 1$ there is no significant difference between our two methods. Finally, we consider the two cases $\alpha = 0.2$ and $\alpha = 0.005$ for three possible δ's, $\delta = 0.5$, $\delta = 1$ and $\delta = 2$. The overall picture is that with $\delta = 2$ the second method is superior to the first, roughly, to the extent that, for $\delta = 0.5$, the first is superior to the second. (Compare the values in the fourth columns of Tables 3.3 and 4.3 with the respective third columns of Tables 3.1 and 4.1.) Also, for $\delta = 2$, naturalness tends to increase the superiority of e_2 over e_1 (Compare the fourth columns of Tables 3.3 and 4.3; for $n \geqslant 4$, the values in 4.3 are bigger.)

The results are shown graphically in figures 1–4. Figure 1 exhibits the standard solution with $\delta = 0.001$; figure 2, the effect of changing (α, β); figure 3, the effect of letting δ vary from 0.05 to 1; and figure 4, the results of inverting δ.

Note that the inversion of δ does not symmetrically reverse the ranking of $e_1(n)$ and $e_2(n)$. Thus, if $\delta = 2$, $e_1(n)$ has still a higher confirmation value for $n = 1, 2$; only for $n \geqslant 4$ does $e_2(n)$ take precedence. As n grows, the picture gains in symmetry: $e_2(n)$ tends to be superior to the same extent that $e_1(n)$ is superior for $\delta = 0.5$. The asymmetry is partly due to the fact that δ is the ratio of ravens to non-black non-ravens, *not* of ravens to non-blacks. Its inversion does not mean a full exchange of 'ravens' and 'non-blacks'; that non-blacks may outnumber the ravens has some positive weight even if $\delta = 2$. In addition, naturalness plays here a role, as is indicated by the fact that the asymmetry is more pronounced if the values of α and β are higher. A precise analysis of this role requires that we take as background knowledge not δ, but the ratio, say δ^*, of ravens to non-blacks. But δ^* and the ratio, say x, of black ravens to ravens cannot be considered as probabilistically independent; for example, $x = 0$ implies $\delta^* \leqslant 1$. This makes the choice of a prior more problematic (it is the reason for my using δ instead of δ^*). But one can prove (i) that for all priors, if $\delta^* \geqslant 1$ then, for *all n*, $e_2(n)$ is not inferior in confirmation value to $e_1(n)$; (ii) that if $\delta > 1$ and if the prior probability that some ravens are black and some are not is not zero, then, for *all n*, $e_2(n)$ is actually superior. We can therefore conclude that – whatever effect of naturalness is expressed in the asymmetry – it has minor weight (see also the postscript after the Notes).

The Hebrew University of Jerusalem

NOTES

[1] No puzzle arises if blackness is considered an *a priori* prerequisite for being a raven. Hence the possibility of a non-black bird possessing properties which will induce us to classify it as a raven is, in principle, not excluded. That all ravens are black is, if true, a true empirical statement, just as the statement that all swans are white is empirically false.

[2] This is a rather rare disease known to affect Jews of East European origin.

[3] Note, however, that 'all ravens to be encountered within the next n years are black' *is* equivalent to 'all non-black objects to be encountered within the next n years are not ravens'. Thus, interpreting 'all A's' as 'all A's to be encountered within the next n years' does not give rise to a "nearly-all" solution.

[4] By backing such a generalization, the argument runs, one loses when a counter-example is found, but in order to win, infinitely many instances must be verified. Hence, one cannot win in any finite time and the bet is irrational at any odds. Indeed, if these be the conditions of winning and losing, no odds are acceptable. Yet this by no means entails that the generalization has probability zero. For under these conditions one is not betting on the truth of our hypothesis, against its falsity, but on its being proved against its being refuted, where 'proof' means no less than the actual verification of all instances. This difficulty does not arise in ordinary betting situations, where, by virtue of background knowledge, one is assured of finding the truth value of the contended hypothesis within reasonable time. For example, the outcome of a race will be known through watching it or by reading the newspaper. In order that the acceptable odds reflect one's estimate of h's probability it is necessary that his odds for 'h is true' against 'h is false' should be the same as his odds for 'h will be proved' against 'h will be refuted'. Consider the following example where the failure of this condition has nothing to do with universal generalizations. Mr Cohen is convinced that Hugo, a deceased official of the Ministry, passed information to a foreign country. Unfortunately, the only way to decide the issue is through the contents of a letter to be found in a certain safe. Cohen believes that Hugo had access to this safe and he reasons that, had he been guilty, he would probably have destroyed the incriminating letter. Thus, if Cohen is wrong the letter would prove it, but if he is right the letter would not exist. The very truth of h causes the evidence to be destroyed! Under these conditions Cohen cannot rationally bet on his hypothesis with odds that correspond to his belief in it. If probabilities are to be illustrated as betting odds, then in Mr Cohen's situation, in the case of universal statements, or in any similar case, one should consider imaginary bets, where the outcome is to be decided through revelation by an omniscient oracle.

[5] Goodman's theory of entrenchment has been proved untenable by Zabludowski (see the exchange in the *Journal of Philosophy*, 1975, numbers 5 and 21). The weakness of Goodman's account is that it makes projectibility a property of the predicate's extension. If this is given up one is still left with Goodman's main claim that entrenchment (which corresponds to what I, following Quine, refer to as "naturalness") is a combination of mere chance and success. Predicates are chosen by accident and they become entrenched by virtue of successful usage. This would imply that instead of the predicates 'conducts' and 'isolates' that figure in electricity theory we could just as well have had 'condulates' and 'isoducts', where the first is 'conducts if the date is before A.D. 2000 and isolates if the date is later' and the second is similarly construed. It is merely by chance that between 1729 and 1750, when conductivity was discovered, we came to use the first pair, not the second. The same would apply to every scientific term that came or will come into usage. An absurd conclusion. Goodman could argue that our choice of 'conducts' is determined by other predicates in usage at that time, which predicates themselves became entrenched according to his account. But the same story would be then repeated and the final upshot is to push back the choice by chance to some primordial semi-mythical past of which very little can be said. Thus,

Goodman's account amounts to claiming that our choice of predicates is unexplainable beyond the fact that it passed the test of success (i.e., leading to successful predictions).

[6] Note that the criterion of survival does not determine a unique class of natural predicates. Ad-hoc artificial constructions such as Goodman's 'grue' could have equally well led to survival, at least – in the case of 'grue' – up to A.D. 2000.

[7] For example, suppose that h_1 and h_2 are two hypotheses, and that, because of the smallness of each $p(h_i)$, it is difficult to estimate the ratio $p(h_1)/p(h_2)$. If e is implied by each hypothesis, then this ratio is also $p(h_1/e)/p(h_2/e)$. So the method consists in looking for an e, such that at least one $p(h_i/e)$ is sufficiently large to be within our intuitive range. Then we ask ourselves: suppose we know e to be true, what would then be the ratio of the two probabilities? The answer is the actual ratio.

[8] Inasmuch as different systems of beliefs may be exhibited within a given community, or even by the same person, we should consider not a single probability function but a certain family. Agreement is reached when all lead to the same verdict. Also, leaving certain details of a function unspecified, amounts, in fact, to having a family of functions – namely, the functions arising out of the different ways of fixing the "free" details.

[9] One is a continuity rule concerning quantifiers, known also in the literature as condition (G). Another is a rule to the effect that consistent sentences should be accorded non-zero probabilities – a condition that can be always satisfied, since the language is countable. However, the price of this second rule, in terms of the complexity of the function satisfying it, is very high; so a restricted version is perhaps more plausible.

[10] For example, $p(h_1) = 10^{-2}$, $p(h_2) = 10^{-15}$, $p(e) = 0.011$, then $p(h_1/e) = 0.91$, $p(h_2/e) = 0.91 \cdot 10^{-13}$. The increase from 0.01 to 0.91 is extremely significant; that from 10^{-15} to $0.91 \cdot 10^{-13}$ is of no importance, although both are by the same factor. The true picture is indicated by the confirmation rate, $\mathrm{Cr}(h, e)$, which for $h = h_1$ has the value 10.89 while for h_2 it is $1 + 9.10^{-14}$. When $p(h_1/e)$ is sufficiently high, h_1 is accepted and, hence, h_2 is rejected. Unlike probability growth which is continuous, acceptance is a matter of a yes–no decision. All this creates the impression that h_2 is in worse condition as a result of verifying e, though h_2 implies e! This is really an optical illusion; the whole story is contained already in the ratio $p(h_1)/p(h_2)$ of *a priori* probabilities.

[11] It is presupposed here that the relative frequency of A's (within an inclusive bigger domain) tells us in itself practically nothing about their absolute number, for either no knowledge is presupposed concerning the absolute size of the bigger domain, or both are regarded as very, very large, or potentially infinite.

[12] I could have assumed to start with that C is the collection of all objects in the world. I prefer to include the possibility of a restricted C in order to have a more realistic experiment. Note that replacing everywhere 'raven' by 'raven in C' and 'object' by 'object in C' would get us the same effect as the assumption that C is the universal class.

[13] Here it is presupposed that the total number of ravens in the world, say $\mathcal{N}$, is very large compared to n – the size of our sample; so that by drawing out n black ravens the ratio of black ravens to ravens is practically unchanged. When every drawn raven has no chance of being redrawn the precise value of $e_1(n)$, given that $w_{BR} = x$, is $(x\mathcal{N})! \cdot (\mathcal{N} - n)!/\mathcal{N}!\,(x\mathcal{N} - n)!$. This value is smaller than x^n, but,

using Stirling's formula, one can show that it is bigger than $x^n \cdot (1 - n \cdot [(1 - x)/(\mathcal{N} - n)]x)^n$. The factor $(1 - n \cdot [(1 - x)/(\mathcal{N} - n)])^n$ is very near 1, unless x is very small. But for small x the probability x^n is negligible and contributes practically nothing to the value of $p(e_1(n))$ which is obtained by integrating over x. The relative error in the value of the integral (where a uniform distribution of x is assumed, except for the end point 0 and 1) is found to be no more than $n/\mathcal{N}$ (i.e., the error percentage is not more than $100n/\mathcal{N}$). Furthermore if α is the initial probability of h then the relative error in our final results is no more than $(n/\mathcal{N})[1 - \alpha/((n - 1)\alpha + 1)]$ which, using any reasonable estimate of $\mathcal{N}$, is negligible. Note that x^n is the exact probability provided that, in the case that C contains all ravens, every drawn and tested raven is returned to the collection before the next drawing. If C contains m ravens each drawn raven is to be returned with probability $m/\mathcal{N}$. We can imagine a random device effecting this procedure.

[14] An observation similar to Note 13 applies here as well. $(1 + \delta(1 - y))^{-n}$ is the exact value if every drawn non-black object is returned to the pool. Otherwise, $p(e_2(n))$ is somewhat smaller but the relative error is negligible. One can show it to be at most $(n/K)[1 - \alpha]/[(n - 1)\delta\alpha + 1]$ where K = number of non-black objects in C.

[15] It can be shown that, in all cases, $p(e_2(n)/C_{RD} = \delta \wedge w_{BR} = x)$ is at least $(1 - \delta(1 - x))^{-n}$ and at most $(1 - \delta(1 - x))^n(1 + (2/l)n^2 \cdot x(x - 1))$ where l is the number of D's in C. In order to find $p(e_2(n)/C_{RD} = \delta)$ I shall integrate over x and then the relative error in the value of the integral is no more than $6/(\delta \cdot l) + 2n/(n \cdot l + \delta^n)$. Since the integral is only one of the terms in the final result, in the rest of which there is no error, the actual relative error is considerably smaller. Also the weight of this integral decreases strongly with n. Note that $\delta \cdot l$ is the number of ravens in C, so if C contains at least 1000 ravens the error percentage is smaller than 1%. Finally, the error is considerably smaller if C contains an appreciable fraction of the world population of ravens, In the extreme case where C contains the whole population there is no error, no matter what the number of ravens is.

[16] That is not to say that, in general, high initial probability values for generalizations constitute by themselves a sufficient mark of naturalness. 'All A's are B's' can be assigned a high value not because of the naturalness of 'A', or of the pair, but by virtue of its logical equivalence to another generalization; 'all non-blacks are not ravens' is assigned the same value as 'all ravens are black'. In general, the status of a predicate can be determined only by its role in a sufficiently wide context. The behaviour of the probability over an extensive class of sentences should be considered. Here I am attempting something much more modest. I consider an isolated situation involving only 'raven' 'black' their negations and sentential composites, where, furthermore, only the pair 'raven' – 'black' is a candidate for naturalness.

[17] This is not to say that for all purposes we can treat $\mathcal{N}$ as if it were infinite. For example, since, in all probability, there are less than 10^{20} ravens, the statement that w_{BR} lies in an open interval of the form $(i \cdot 10^{-20}, (i + 1)10^{-20})$ should be assigned value 0. Statements of this form do not concern us here, neither do they as a rule concern the statistician. Those in which we are interested, namely that all ravens, or all ravens in C, are black, are practically unaffected by the switch to continuous distributions.

[18] This includes all density functions which, in some neighborhood of 1 the form $(1 - \varepsilon, 1]$, where $\varepsilon > 0$, are continuous, bounded from above and bounded from

below by some positive number. One can also consider priors that give positive probabilities to finitely many distinguished points (besides the extremes 0 and 1). The last criterion holds for these as well. Conditions can also be given for priors that give non-zero probabilities to sets of Lebesque measure 0.

POSTSCRIPT

(1) The effect of naturalness that has been pointed out at the end of §5 corresponds to the following intuition: the more natural the pair 'raven'–'black' is, the more certain we are that blackness is uniform with respect to ravens, i.e., that either all ravens are black or none are. But taking this disjunction as a premise, a single black raven suffices to prove that all are black, whereas a single non-black non-raven does not. Why is it that such a prima facie clear intuition is expressed in our analysis as a relatively small quantitative effect? Actually the intuition is misleading. It is another instance of our tendency to emphasize simple regularities and causal relationships and to grossly underestimate frequency effects. For instance, if δ^* (the known ratio of A's to non-B's in our sampling population) is 1, then h (the hypothesis that all A's are B's) is confirmed by $e_1(n)$ (n A's that are B's) exactly to the same extent that it is confirmed by $e_2(n)$ (n non-B's that are non-A's). In particular if we presuppose complete uniformity of B in A (i.e. that either all A's or none are B's) then from $\delta^* = 1$ and from the existence of a non-B which is a non-A our h logically follows – a fact which is not appreciated at first glance. If $\delta^* < 1$, then $e_1(n)$ is superior; but the difference between the two diminishes to the extent that δ^* is near 1. On the whole, the value of δ^* plays the central role even when complete uniformity is presupposed. (In which case $p(h/e_1(n)) = 1$ and $p(h/e_2(n)) = (1 + (1 - \delta^*)^n \cdot (1 - \alpha)/\alpha)^{-1}$, where α is the prior probability of h; hence $(1 - \alpha)/\alpha$ also plays an important role provided that n is not too large. Note that in this case $\mathrm{Cr}(h, e)$ is infinite; hence, as remarked in §3, the ratio of Cr's is not a good indicator of the relative confirmation value. Some other function, say the difference of the $p(h/e_i(n))$, or their ratio, should be used). Presupposing complete uniformity is justified if accumulated evidence, or existing theory – itself highly supported – has made it practically certain. From the behaviour of a sample of a chemically defined substance, say its solubility in water or its melting point, we deduce the behaviour of all samples of the given substance. But this paper is about the different case where complete uniformity is still quite uncertain. The expectations of uniformity that concern us here derive not from an advanced theory, but from some initial dispositions to classify things in certain ways when we are still at the stage of looking for evidence. In this case a prior probability of 0.6 that either all or none of the A's (ravens) are B's (black) marks quite a high initial expectation of regularity, i.e. a high degree of naturalness of the predicates. The deviation from presupposed uniformity adds, on the whole, a considerable weight to the frequency effect.

There is little point in trying to analyse the effects with greater accuracy than that of §5. The finer details will depend on finer features of the prior. But in fact we do not have a specific prior but a family of reasonable candidates that share broad characteristics. Hempel's puzzle has to do with non-quantified intuitions. An analysis that brings out the main factors and yields a rough idea of their relative importance is the best we can achieve.

(2) Though I do not survey the literature I feel obliged to refer to Janina Hosiasson-Lindenbaum's 1940 paper 'On confirmation', *Journal of Symbolic Logic* **5**. That paper is the first to carry out a thorough probabilistic analysis of the paradox, as well as of other aspects in intuitive confirmation theory. The notation and the mathematical details might have deterred philosophers from studying it closely, but the analysis is impeccable and, to my mind, surpasses much later work in this direction; it ought to mark a milestone in confirmation theory. Hosiasson works in a probabilistic system that assigns numerical values to pairs of sentences, her $c(h, e)$ corresponds to my $p(h/e)$. Her results are of a general character and apply to all priors. She has spotted both the effect of frequency and that of naturalness that is discussed in (1) of this postscript (she refers to it as the c-value of homogeneity). But she did not try to combine them and find their relative weight.

I should also like to mention J. L. Mackie's 'The paradox of confirmation', *British Journal for the Philosophy of Science* **13** (1962–63), in which he sums up neatly the role of background knowledge, the meaning of "observation of a black raven" and "observation of a white show" and shows that various Popperian arguments are either mistaken or are corollaries of probabilistic principles.

(3) Concerning our little story of Cohen and Hugo in note 4, it is, of course, to be understood that the absence of the letter would not, in itself, constitute any appreciable confirmation of Hugo's treason. People other than Cohen either do not know of the letter, or may explain its absence in many ways without presuming Hugo's guilt. (The effect of this absence on Cohen's beliefs is irrelevant. But even here we may suppose that, convinced as he already is of Hugo's guilt, no appreciable change will result.) The reader can, no doubt, make up similar stories that establish the same point.

(4) Professor Goodman has pointed out to me that there has been a sequel to the Goodman–Ulian versus Zabludowski exchange, namely the following articles, all in the *Journal of Philosophy*: 'Projectibility unscathed' by Goodman and Ulian, 1976; Zabludowski's rejoinder 'Quod periit, periit', 1977; and a rejoinder to the rejoinder, 'The short of it', 1978. Having gone through all these I see no reason to change my conclusion that Goodman's theory is untenable (see note 5). The point that Goodman and Ulian arrived at after long and tortuous arguments, with many revisions and turnings, is that in order to save the theory they must render it highly counter-intuitive and, moreover, uninformative. On their account a hypothesis such as "all emeralds are green" can be projected only after we have already accepted another, more complex hypothesis, namely: "every emerald is either green and fusible or not green and not fusible" and "every emerald is either green and gravitates or is not green and does not gravitate", and so on ad infinitum. Goodman's theory can derive the projectibility of our original hypothesis only if it takes as an underived axiom the acceptance of all those in the second bunch.

TABLES

Values are expressed to 3 decimal places; e.g., '1.000' indicates a value which differs from 1 by no more than 5×10^{-4}. Cr is the confirmation rate. '$\mathrm{Cr}(h, e_i)$' stands for '$\mathrm{Cr}_\delta(h, e_i(N))$', similarly for '$p(h/e_i)$'.

TABLE 1.1

$\alpha = 0.05$; $\beta = 0.1$; $\delta = 0.001$					
n	$p(h/e_1)$	$\mathrm{Cr}(h, e_1)$	$p(h/e_2)$	$\mathrm{Cr}(h, e_2)$	$\dfrac{\mathrm{Cr}(h, e_1)}{\mathrm{Cr}(h, e_2)}$
1	0.105	1.064	0.050	1.000	1.062
2	0.180	1.118	0.050	1.000	1.118
4	0.227	1.229	0.050	1.000	1.229
8	0.346	1.453	0.050	1.000	1.453
16	0.500	1.900	0.050	1.000	1.899
32	0.660	2.794	0.051	1.001	2.792
64	0.793	4.582	0.052	1.002	4.574
128	0.884	8.159	0.053	1.004	8.129
256	0.938	15.312	1.057	1.007	15.990
512	0.968	29.618	0.064	1.015	29.166
1024	0.984	58.229	0.081	1.034	56.335

TABLE 1.2

$\alpha = 0.1$; $\beta = 0.2$; $\delta = 0.001$					
n	$p(h/e_1)$	$\mathrm{Cr}(h, e_1)$	$p(h/e_2)$	$\mathrm{Cr}(h, e_2)$	$\frac{\mathrm{Cr}(h, e_1)}{\mathrm{Cr}(h, e_2)}$
1	0.222	1.157	0.100	1.000	1.157
2	0.300	1.286	0.100	1.000	1.286
4	0.418	1.543	0.100	1.000	1.542
8	0.563	2.057	0.100	1.000	2.056
16	0.708	3.086	0.101	1.001	3.083
32	0.825	5.143	0.102	1.002	5.133
64	0.903	9.257	0.104	1.004	9.221
128	0.949	17.486	0.107	1.008	17.346
256	0.974	33.940	0.117	1.017	33.392
512	0.987	66.860	0.130	1.035	64.610
1024	0.993	132.690	0.164	1.076	123.260

TABLE 1.3

$\alpha = 0.2$; $\beta = 0.4$; $\delta = 0.001$					
n	$p(h/e_1)$	$\mathrm{Cr}(h, e_1)$	$p(h/e_2)$	$\mathrm{Cr}(h, e_2)$	$\frac{\mathrm{Cr}(h, e_1)}{\mathrm{Cr}(h, e_2)}$
1	0.500	1.600	0.200	1.000	1.600
2	0.600	2.000	0.200	1.000	2.000
4	0.714	2.800	0.210	1.001	2.798
8	0.818	4.400	0.201	1.001	4.395
16	0.895	7.600	0.202	1.002	7.582
32	0.943	14.000	0.204	1.005	13.932
64	0.970	26.800	0.208	1.010	26.540
128	0.985	52.400	0.216	1.020	51.375
256	0.992	103.600	0.232	1.041	99.475
512	0.996	206.000	0.266	1.082	189.094
1024	0.998	410.800	0.337	1.206	340.636

TABLES 2.1–2.4

$\alpha = 0.1$; $\beta = 0.2$

TABLE 2.1 $\delta = 0.05$

n	$p(h/e_1)$	$Cr(h, e_2)$	$\frac{Cr(h, e_1)}{Cr(h, e_2)}$
1	0.103	1.003	1.154
2	0.105	1.006	1.278
4	0.111	1.013	1.524
8	0.123	1.026	2.005
16	0.148	1.056	2.921
32	0.202	1.128	4.558
64	0.312	1.308	7.079
128	0.475	1.715	10.194
256	0.646	2.539	13.367
512	0.785	4.185	15.975
1024	0.880	7.476	17.747

TABLE 2.2 $\delta = 0.1$

n	$p(h/e_2)$	$Cr(h, e_2)$	$\frac{Cr(h, e_1)}{Cr(h, e_2)}$
1	0.105	1.006	1.150
2	0.111	1.012	1.270
4	0.132	1.026	1.504
8	0.147	1.055	1.950
16	0.201	1.126	2.741
32	0.309	1.303	3.948
64	0.473	1.709	5.418
128	0.645	2.533	6.904
256	0.785	4.179	8.123
512	0.880	7.470	8.950
1024	0.936	14.053	9.442

TABLE 2.3 $\delta = 0.5$

n	$p(h/e_2)$	$Cr(h, e_2)$	$\frac{Cr(h, e_1)}{Cr(h, e_2)}$
1	0.125	1.028	1.125
2	0.153	1.062	1.211
4	0.214	1.145	1.348
8	0.338	1.359	1.514
16	0.517	1.863	1.665
32	0.689	2.893	1.778
64	0.818	4.950	1.870
128	0.901	9.064	1.929
256	0.948	17.239	1.963
512	0.973	33.750	1.981
1024	0.986	66.660	1.990

TABLE 2.4 $\delta = 1$

n	$p(h/e_2)$	$Cr(h, e_2)$	$\frac{Cr(h, e_1)}{Cr(h, e_2)}$
1	0.146	1.054	1.098
2	0.200	1.125	1.143
4	0.316	1.315	1.173.
8	0.500	1.800	1.143
16	0.682	2.829	1.091
32	0.816	4.882	1.053
64	0.900	9.000	1.029
128	0.948	17.229	1.015
256	0.973	33.686	1.008
512	0.986	66.600	1.004
1024	0.993	132.429	1.002

TABLE 3.1–3.3

$\alpha = 0.05$; $\beta = 0.10$

TABLE 3.1 $\delta = 0.5$

n	$p(h/e_2)$	$\mathrm{Cr}(h, e_2)$	$\frac{\mathrm{Cr}(h, e_1)}{\mathrm{Cr}(h, e_2)}$
1	0.062	1.013	1.048
2	0.076	1.028	1.087
4	0.107	1.063	1.156
8	0.177	1.154	1.259
16	0.306	1.370	1.387
32	0.477	1.816	1.538
64	0.649	2.710	1.691
128	0.789	4.499	1.814
256	0.882	8.075	1.896
512	0.938	15.228	1.945
1024	0.968	29.534	1.972

TABLE 3.2 $\delta = 1$

n	$p(h/e_2)$	$\mathrm{Cr}(h, e_2)$	$\frac{\mathrm{Cr}(h, e_1)}{\mathrm{Cr}(h, e_2)}$
1	0.073	1.024	1.037
2	0.100	1.056	10.59
4	0.164	1.137	1.081
8	0.293	1.343	1.082
16	0.469	1.788	1.062
32	0.646	2.682	1.042
64	0.788	4.471	1.025
128	0.882	8.077	1.014
256	0.938	15.200	1.007
512	0.968	29.506	1.004
1024	0.984	58.118	1.002

TABLE 3.3 $\delta = 2$

n	$p(h/e_2)$	$\mathrm{Cr}(h, e_2)$	$\frac{\mathrm{Cr}(h, e_1)}{\mathrm{Cr}(h, e_2)}$	$\frac{\mathrm{Cr}(h, e_2)}{\mathrm{Cr}(h, e_1)}$
1	0.091	1.045	1.016	0.984
2	0.145	1.111	1.006	0.994
4	0.266	1.295	0.949	1.053
8	0.452	1.733	0.839	1.192
16	0.638	2.626	0.723	1.382
32	0.785	4.415	0.633	1.580
64	0.881	7.991	0.573	1.744
128	0.975	15.144	0.539	1.856
256	0.968	25.450	0.520	1.923
512	0.984	58.062	0.510	1.960
1024	0.992	115.285	0.505	1.980

TABLES 4.1–4.3

$\alpha = 0.2$; $\beta = 0.4$

TABLE 4.1 $\delta = 0.5$

n	$p(h/e_2)$	$\mathrm{Cr}(h, e_2)$	$\frac{\mathrm{Cr}(h, e_1)}{\mathrm{Cr}(h, e_2)}$
1	0.253	1.071	1.494
2	0.310	1.160	1.724
4	0.429	1.400	2.000
8	0.619	2.099	2.097
16	0.788	3.773	2.104
32	0.886	7.000	2.000
64	0.940	13.400	2.000
128	0.969	26.200	2.000
256	0.985	51.800	2.000
512	0.992	103.000	2.000
1024	0.996	205.400	2.000

TABLE 4.2 $\delta = 1$

n	$p(h/e_2)$	$\mathrm{Cr}(h, e_2)$	$\frac{\mathrm{Cr}(h, e_1)}{\mathrm{Cr}(h, e_2)}$
1	0 295	1.135	1.409
2	0.400	1.333	1.500
4	0.585	1.929	1.451
8	0.774	3.546	1.241
16	0.882	6.779	1.118
32	0.939	13.200	1.061
64	0.969	26.000	1.031
128	0.984	52.600	1.016
256	0.992	102.800	1.008
512	0.996	205.200	1.004
1024	0.998	410.000	1.002

TABLE 4.3 $\delta = 2$

n	$p(h/e_2)$	$\mathrm{Cr}(h, e_2)$	$\frac{\mathrm{Cr}(h, e_1)}{\mathrm{Cr}(h, e_2)}$	$\frac{\mathrm{Cr}(h, e_2)}{\mathrm{Cr}(h, e_1)}$
1	0.362	1.253	1.277	0.784
2	0.529	1.700	1.178	0.850
4	0.743	3.114	0.899	1.112
8	0.875	6.391	0.689	1.451
16	0.937	12.800	0.594	1.684
32	0.969	25.600	0.547	1.828
64	0.984	57.200	0.523	1.912
128	0.992	102.400	0.512	1.953
256	0.996	204.800	0.506	1.976
512	0.998	409.600	0.503	1.988
1024	0.999	819.200	0.501	1.996

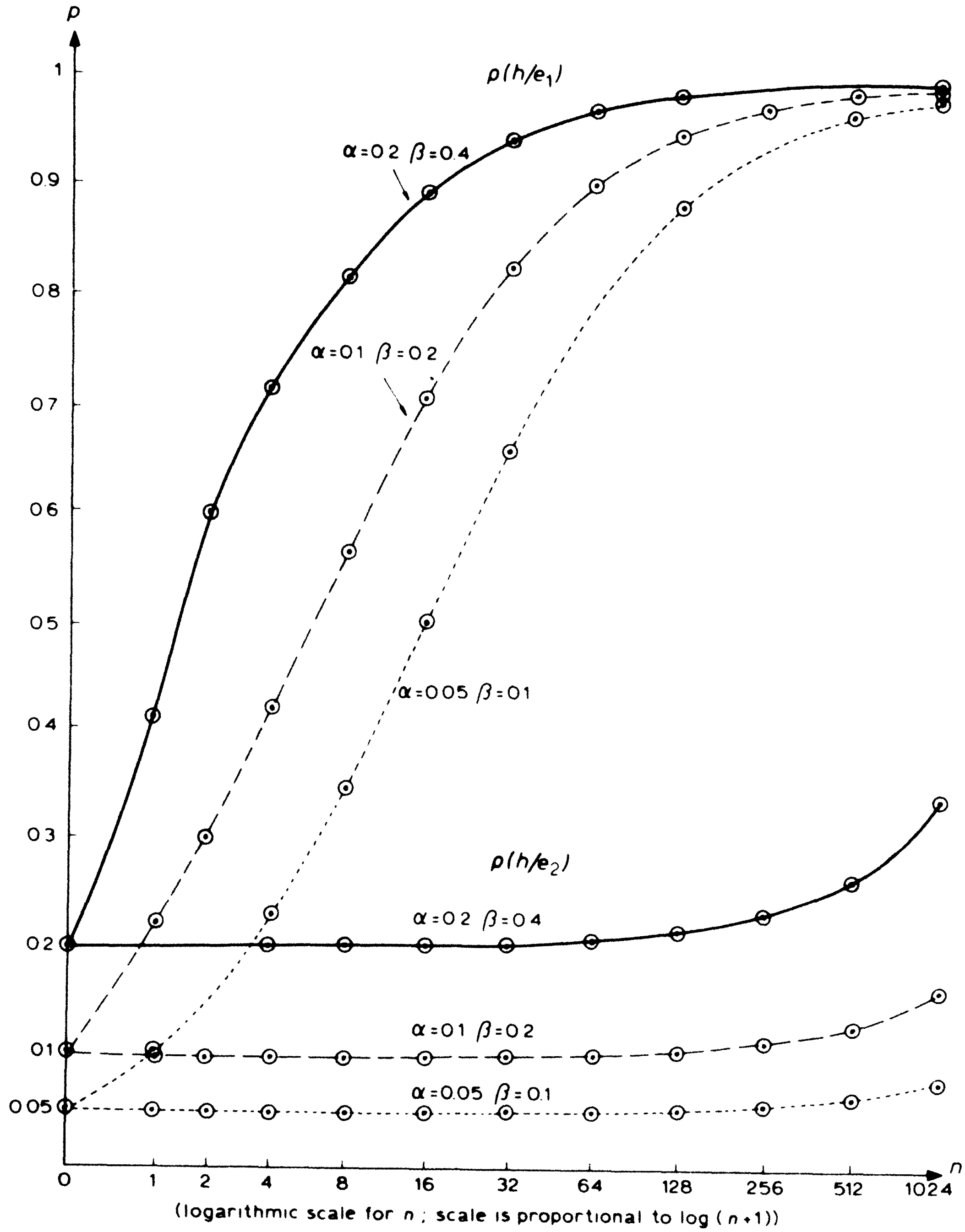

Fig. 1. Solution of Hempel's puzzle (Tables 1.1–1.3). e_1 is relevant evidence; e_2 is not. $\delta = 0.001$.

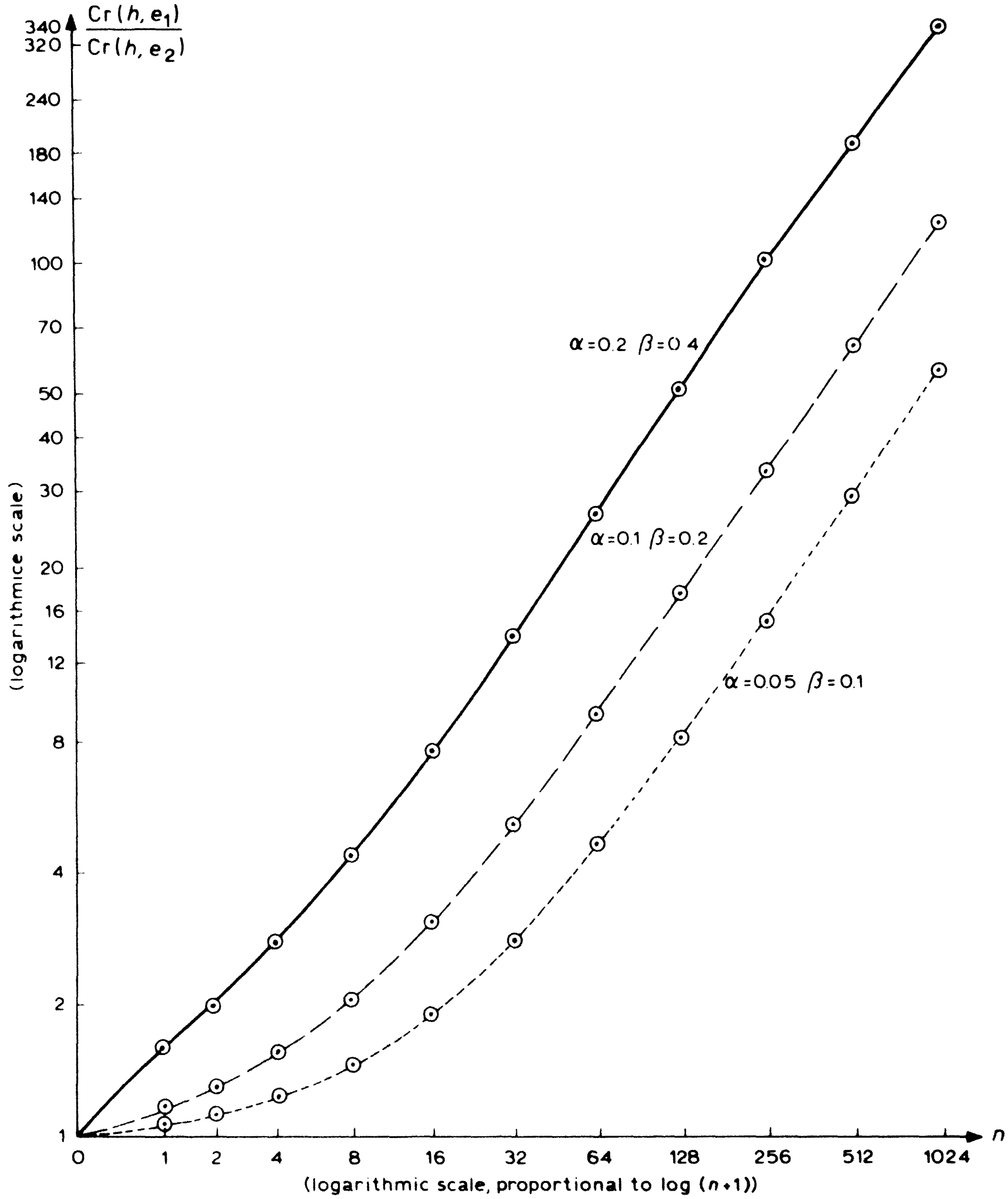

Fig. 2. The effect of naturalness (Tables 1.1–1.3). $\delta = 0.001$.

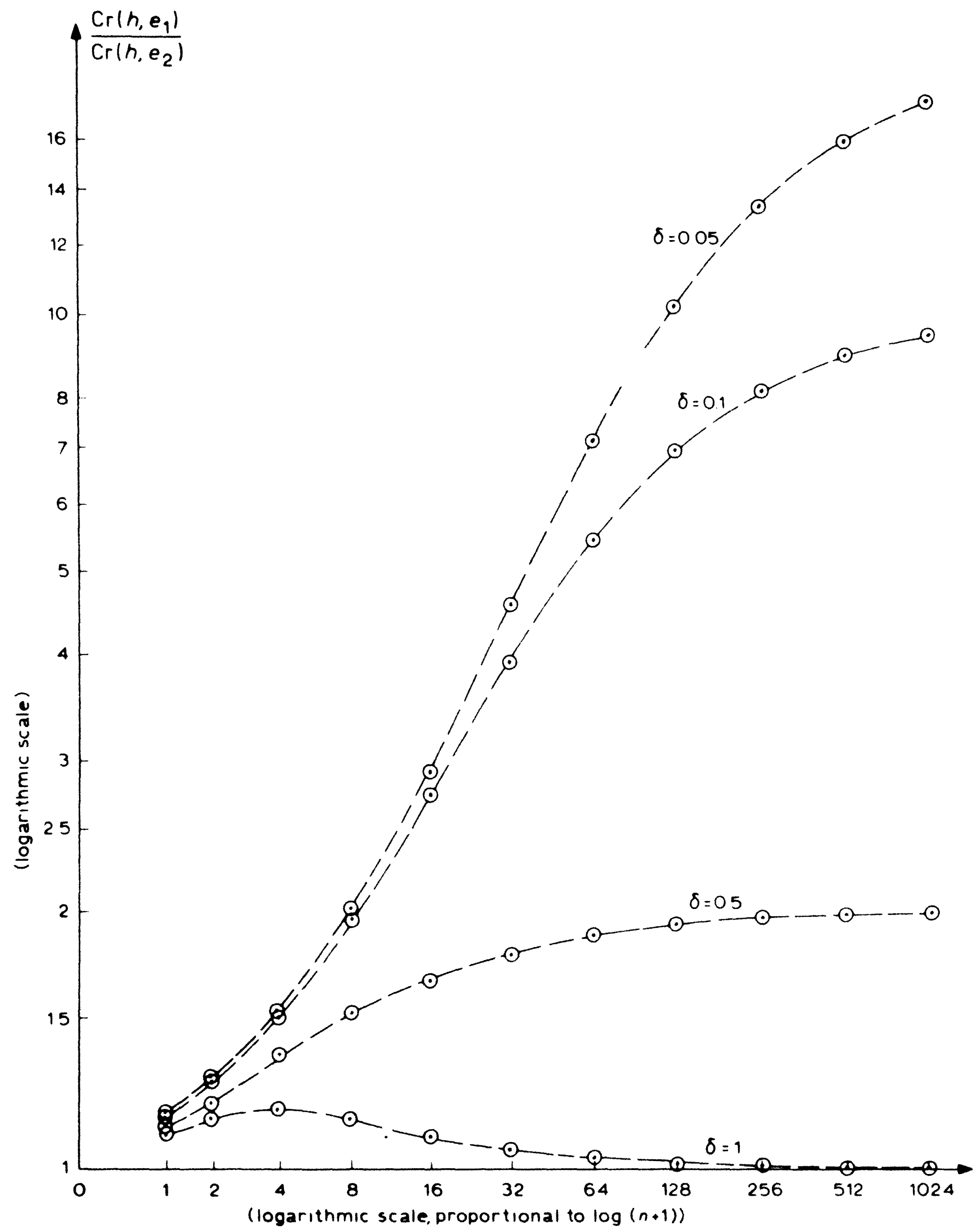

Fig. 3. The frequency effect (Tables 2.1–2.4). As δ tends to 1, the confirmation value of e_2 comes nearer to the confirmation value of e_1. For $\delta = 1$, there is no appreciable difference; $\alpha = 0.1$; $\beta = 0.2$; $1/\delta$ is the asymptotic value.

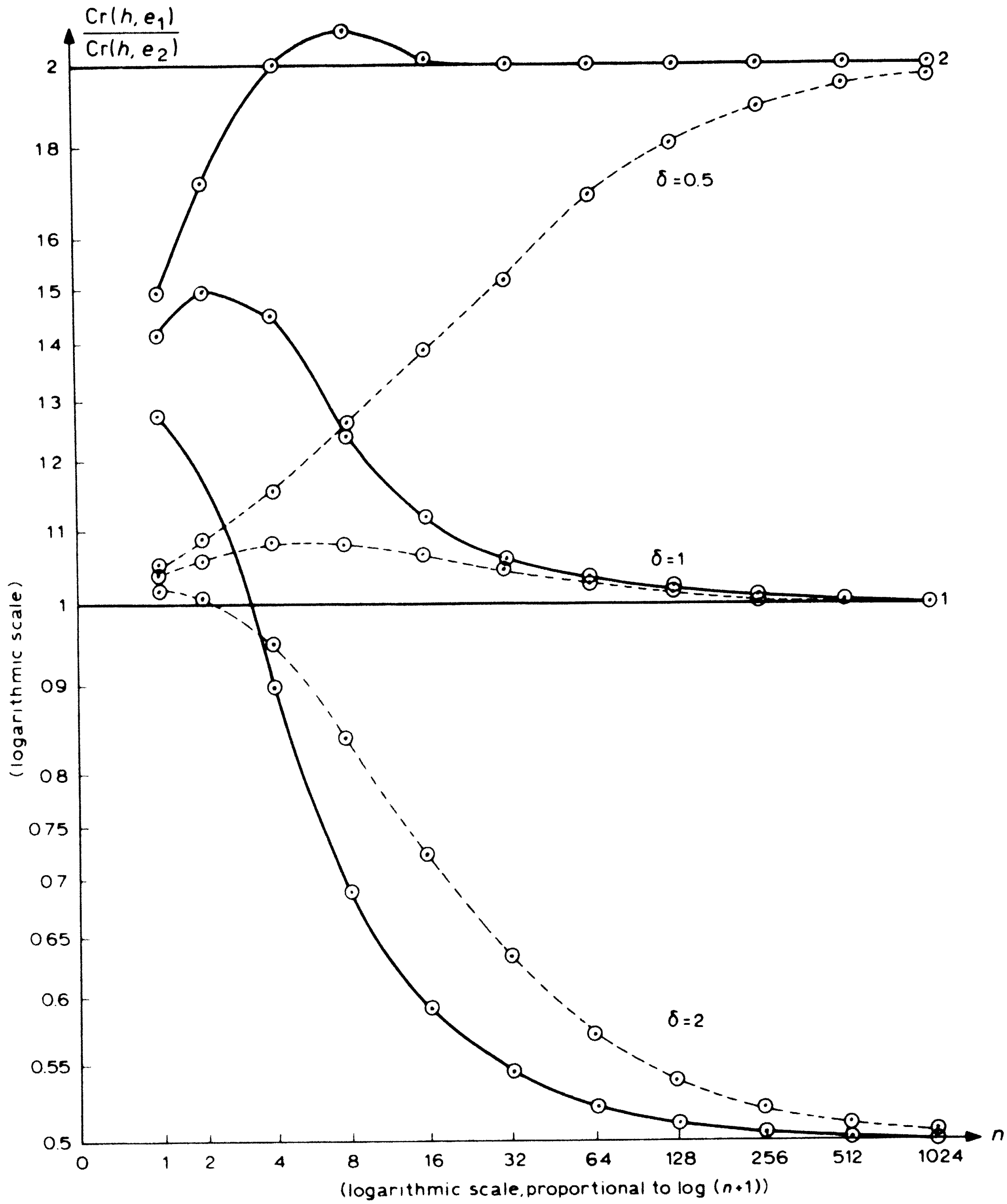

Fig. 4. Frequency reversal (Tables 3.1–3.3; 4.1–4.3). For $\delta = 2$, the confirmation value of e_1 is less than the value of e_2 and the situation is roughly reversed. —— indicates $\alpha = 0.2$; $\beta = 0.4$ (highly natural predicates). - - - indicates $\alpha = 0.05$; $\beta = 0.1$.

Manuscript received 22 June 1978