

[Colin Howson](#)

Does information inform confirmation?

**Article (Accepted version)
(Refereed)**

Original citation: Howson, Colin (2015) *Does information inform confirmation?* [Synthese](#) . pp. 1-15.
ISSN 0039-7857

DOI: [10.1007/s11229-015-0918-7](https://doi.org/10.1007/s11229-015-0918-7)

© 2015 [Springer Science+Business Media Dordrecht](#)

This version available at: <http://eprints.lse.ac.uk/64398/>

Available in LSE Research Online: November 2015

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.

This document is the author's final accepted version of the journal article. There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

Does information inform confirmation?

Abstract

In a recent survey of the literature on the relation between information and confirmation, Crupi and Tentori (2014) claim that the former is a fruitful source of insight into the latter, with two well-known measures of confirmation being definable purely information-theoretically. I argue that of the two explicata of semantic information (due originally to Bar Hillel and Carnap) which are considered by the authors, the one generating a popular Bayesian confirmation measure is a defective measure of information, while the other, although an admissible measure of information, generates a defective measure of confirmation. Some results are proved about the representation of measures on consequence-classes.

Key words: confirmation, information, explication, consequences.

1. Introduction.

Originating famously with Francis Bacon, the pursuit of a provably or even probably reliable set of rules of inductive inference suffered a severe reverse in the mid-eighteenth century, under the onslaught of Hume's celebrated sceptical arguments whose continuing failure to be definitively refuted after two centuries of strenuous attempts was famously dubbed by C.D. Broad 'the scandal of philosophy'¹. And so matters stood until, under the pioneering leadership of Rudolf Carnap, attention was diverted away from trying to refute Hume's arguments² to the less intimidating task of 'explicating' inductive *concepts*,

¹ 1926, p. 67.

² These arguments are all variations on the following: that any attempt to infer anything about the future from the past will necessarily beg the question. Howson (2013) argues that this argument translates

like *degree of confirmation*, *confirming instance*, *evidential support*, and so on, in so doing begetting the thriving enterprise now known as *confirmation theory*.

One of the more interesting sub-projects to emerge from this programme has been an appeal to information-theoretic ideas in the hope that they would illuminate confirmation-theoretic ones. The idea that the acquisition of new information should be related to confirmation and disconfirmation seems on the face of it natural enough: as the authors of a recent extensive survey put it,

the impact of a piece of evidence (data, premise) on a given hypothesis (theory, conclusion) must reflect how the former affects an antecedent state of information concerning the latter. Relatedly, a rational agent would gather evidence because it provides information concerning certain possible states of affairs, i.e., for it can confirm/disconfirm relevant hypotheses. (Crupi and Tentori, 2014, p.81.)

The subject-matter of confirmation theory is however typically *statements*, statements of hypotheses and evidence, and the application of information-theoretic notions required a suitable redefinition of the hitherto purely statistical definition of information introduced in the work of Claude Shannon. That transformation, into what he called *semantic information*, carried by the sentences of some appropriate language L, was undertaken by Carnap in a pioneering paper with Bar Hillel (Bar Hillel and Carnap 1953³), laying the foundations for nearly all subsequent work in the field.⁴

2. Improbability, measure and information

The seminal nature of Bar Hillel and Carnap's paper is duly acknowledged by Crupi and Tentori in their survey, citing its 'canonical formulation' of the basic intuition concerning the amount of information conveyed by a statement:

straightforwardly into a theorem of the probability calculus, and hence that attempts to challenge it while remaining within a probabilistic framework are doomed to failure.

³ An earlier technical report by the two authors (Carnap and Bar Hillel 1952), of which their 1953 paper is a more concise published version, is useful for its additional explanatory remarks and will be cited often in the present paper.

⁴ Carnap's ideas were further developed in a number of influential papers by Hintikka (see e.g. Hintikka 1968, 1970).

Being told that the outcome of the draw is a picture of hearts provides more information in comparison, because this singles out a small subset of possibilities that was initially rather improbable. In philosophy, this basic idea found its canonical formulation in seminal work by Bar-Hillel and Carnap (2014, p.81)

Two ideas are brought together here: the *extent* of the subset of possibilities admitted, and its *probability*. Let $M(a)$ (my notation) denote the class of possibilities admitted by a proposition a in some suitable language L .⁵ Then the complement $*M(a)$ of $M(a)$ is the class of possibilities prohibited by a . The more extensive it is, the more possibilities a 's truth rules out, and the more a can be taken as asserting: indeed, it is easy to see that a logically entails b just in case $*M(b)$ is (set-theoretically) included in $*M(a)$. Accordingly Bar Hillel and Carnap took $*M(a)$ to represent qualitatively the *content* of a .⁶ Their next step was to put a numerical measure on $*M(a)$, and here they were able to exploit Carnap's monumental earlier work on inductive probability (1950) in which he had defined a restricted class of normalised measure functions $m(a)$ on the sentences of L . These functions induce a corresponding class of measures on the $M(a)$, each of which, since $*M(a) = M(\sim a)$, takes the value $m(\sim a) = 1 - m(a)$ on $*M(a)$.

The m -functions are formally probability functions, to signal which I will henceforth write $p(a)$ instead of $m(a)$. $1 - p(a)$ is therefore one explication of the *amount of information in a* . But there is another sense of the amount of information in a for which $1 - p(a)$ is not an appropriate measure. This is the sense in which the informativeness of a sentence can be gauged by how *unexpected* its truth would be in the light of what else you know. Given that $1 - p(a)$ is a decreasing function of probability, it might be thought that this sense is automatically captured by that measure. But this is not so. In the case of two probabilistically independent sentences the amount of information in their conjunction is plausibly the sum of the amounts in each, but to deliver this feature the information-measure, considered as a decreasing function of probability, must be minus the logarithm to some base (greater than 1) of $p(a)$. So now we have *two* probability-

⁵ In Bar Hillel and Carnap's paper $M(a)$ is finite for each a : it is the set of state-descriptions admitted by a when a is formulated in a finite monadic predicate language without equality (informally speaking, the state-descriptions are the finest partition of possibility definable in the language).

⁶ Actually they called it the Content of a (upper case 'c').

based information-measures, $1-p$ and $-\log p$, explicating these two different senses of information, respectively referred to by Bar Hillel and Carnap as *cont* (for 'content') and *inf*.⁷

3. Four explicata and a difference

While these explicata have become canonical, the simple logical framework and the restricted class of probability functions employed by Bar Hillel and Carnap have since largely been abandoned, for a combination of mathematical and philosophical reasons too extensive to go into here, replaced by a largely formalism-free environment where the only probability function involved is some coherent representation of belief, i.e. a finitely or possibly countably-additive probability function defined on at the outset or extendable to a Boolean algebra of propositions. This is the sort of 'broadly Bayesian' context in which Crupi and Tentori locate their own discussion⁸, retaining Bar Hillel and Carnap's explicata *cont* and *inf* but employing their own notation for them as '*inf_D*' and '*inf_R*' respectively⁹.

Bar Hillel and Carnap had also defined two measures of *relative information*, $\text{cont}(a|b)$ and $\text{inf}(a|b)$, based on the corresponding one-place functions. Where 'T' stands for either 'cont' or 'inf', Bar Hillel and Carnap's two-place functions are explicitly defined in terms of the one-place functions as follows:

$$I(a|b) = I(a\&b) - I(b). \quad (1)$$

From (1) we infer that $\text{inf}(a|b) = -\log p(a|b)$ and $\text{cont}(a|b) = 1 - p(b \rightarrow a)$, where ' \rightarrow '

⁷ This is admittedly a very brisk and nuance-free summary. The reader is recommended to consult Bar Hillel and Carnap's own exposition.

⁸ 2014, p.81. It will sometimes be useful in what follows to regard the algebra of propositions as the so-called Lindenbaum sentence algebra of a language L, where logically equivalent sentences are in effect identified (thus there will be just one tautology, T, and one contradiction \perp). But there are many contexts where standard finitary languages are incapable of expressing the propositions of interest: for example, the infinite disjunction 'a ticket numbered 1 will win or a ticket numbered 2 will win or a ticket numbered 3 will win or', in the context of a countably infinite lottery whose tickets are labelled by the (numerals for) the natural numbers.

⁹ 2014, p.84. The 'R' is for 'ratio', the 'D' for 'difference'; we shall see why in the following section.

is material implication. (1) suggests an interpretation of $I(a|b)$ as a measure of excess, or added, information. Indeed, turning it round, we have $I(a&b) = I(a|b) + I(b)$, i.e. adding $I(a|b)$ to the information in b gives the information resulting from adding (conjoining) a to b . Bar Hillel and Carnap put it thus: $I(a|b)$ is 'the relative (or additional or excess) information' of a with respect to b (1952, p.8¹⁰).

(1) generates interesting notions of informational independence for each of inf and cont . We can say that a is informationally independent of b just in case $I(a|b) = I(a)$, and hence just in case $I(a&b) = I(a) + I(b)$, which sounds intuitively right and is symmetrical in a and b . Instantiating this with inf , we have $-\log p(a&b) = -[\log p(b) + \log p(a)]$, i.e. $p(a&b) = p(a)p(b)$, so informational independence holds for inf just when a and b are probabilistically independent. Cont is dual to probability in that the same formulas involving $\&$ and \vee symmetrically hold for both, from which follows the analogue of the general addition rule for probabilities:

$$\text{cont}(a) + \text{cont}(b) = \text{cont}(a&b) + \text{cont}(avb). \quad (2)$$

Hence informational independence holds iff $\text{cont}(a&b) = \text{cont}(a) + \text{cont}(b)$ iff $\text{cont}(avb) = 0$ iff $p(avb) = 1$: i.e. the measure of their contingent content in common is zero. So informational independence holds for cont just when a and b are content-disjoint.

Crupi and Tentori also define a corresponding pair of relative functions, $\text{inf}_D(a|b)$ and $\text{inf}_R(a|b)$. While they do not say explicitly what 'presystematic' notions of relative information these are intended to capture, it is clear from their text that each function is simply the result of substituting the updated probability $p(a|b)$ for $p(a)$ in the measures $1 - p(a)$ and $-\log p(a)$ respectively; i.e. $\text{inf}_R(a|b) = -\log p(a|b)$ and $\text{inf}_D(a|b) = 1 - p(a|b)$ ¹¹. $\text{inf}_R(a|b)$ is extensionally the same function of a and b as Bar Hillel and Carnap's $\text{inf}(a|b)$, but not so $\text{inf}_D(a|b)$ and $\text{cont}(a|b)$, which are easily seen to differ from each other by a factor $p(b)$. A result of Milne (2014, cited by Crupi and Tentori) implies that the functions $\text{inf}_D(a|b)$ and $\text{inf}_R(a|b)$ can be interpreted as measures of information-added, raising the question of how in that case $\text{inf}_D(a|b)$ and Bar Hillel and Carnap's $\text{cont}(a|b)$

¹⁰ Bar Hillel and Carnap allow b to be a class of statements, like a body of background assumptions, but if, as is usually assumed, the class is finite it can of course be represented by a single conjunction

¹¹ Thus conditionalisation is implicitly assumed.

can consistently differ. The answer is that two different notions of addition are being appealed to: for Bar Hillel and Carnap it is arithmetical addition expressed by (1), whereas in Milne's paper it is something like proportional addition: 'a proposition adds the more information to b the greater the proportion of possibilities left open by the corpus b that it rules out' (2014, p.253).¹²

While that aspect of Crupi and Tentori's relative measures has some technical interest, their own interpretation of them as merely the result of updating the probability function in each case not only identifies an important concept left unnoticed, or 'unexplicated', by Bar Hillel and Carnap, but is also indispensable in delivering the connection Crupi and Tentori aim to establish between information and confirmation. As they put it,

If the evidence acquired decreases (increases) the degree of unexpectedness of a hypothesis of interest, the credibility of such hypothesis is thereby positively (negatively) affected. A simple way to convey this natural idea is to represent the belief change concerning h provided by e , $bc(h,e)$, by means of the plain difference between $\text{inf}(h)$ and $\text{inf}(h|e)$. [On substituting respectively inf_R and inf_D] two classical confirmation measures are thus immediately recovered:

$$bc_R(h,e) = \text{inf}_R(h) - \text{inf}_R(h|e) = \log[p(h|e)/p(h)]$$

$$bc_D(h,e) = \text{inf}_D(h) - \text{inf}_D(h|e) = p(h|e) - p(h)$$

(2014, p.82)

These equations certainly look like a neat demonstration of some of the 'key connections between information and confirmation' claimed to exist by Crupi and Tentori (2014, p.81). Despite that apparently happy concordance, my own view of the matter, which I will argue in what follows, is very different from theirs, and my conclusion will be that

¹² Milne shows that a set of postulates based on this heuristic principle implies that the information added to b by a , in his sense, is a decreasing function f of $p(a|b)$ with $f(1) = 0$, and he points out that both $-\log p(a|b)$ and $1-p(a|b)$ qualify in this role. His defence against the charge that the notion of a proportion in infinite possibility-spaces, where no limiting procedure exists, is meaningless is that

a probability distribution P is a (weighted) measure of possibilities and so, with a measure of possibilities in place, we restore propriety by saying that a proposition adds the more information to b (according to the distribution P) the smaller $P(a \wedge b)/P(b)$, i.e., the smaller $P(a|b)$. (*ibid.*)

This defence puts more weight on 'weight' than it should probably be asked to bear: a Bayesian probability, for example, is just a coherent degree of belief having nothing to do with 'proportions' of possibilities.

despite their optimistic assessment, the connections are superficial at best and certainly far from 'key'. That is to come; first, we need to review some salient facts about bc_R and bc_D (I will use Crupi and Tentori's terminology).

4. Information and confirmation.

bc_D , i.e. $p(h|e) - p(h)$, is of course a favourite Bayesian measure of evidential support, with several independent considerations in its favour. One of these is the simple naturalness of subtracting the prior from the posterior probability. Another, as is well-known and easily checked, is that any consequence e of a hypothesis h confirms h so long as $0 < p(e), p(h) < 1$. Though this might seem to invite the so-called tacking paradox, since the conjunction of h with any, even nonsensical, statement h' will also be positively supported, the paradox is neatly avoided by another easily-checked fact, that $bc_D(h \& h' | e) = p(h' | h) bc_D(h | e)$. So if h' is tacked on to h , the support of the conjunction $h \& h'$ measured by bc_D will be reduced in proportion to how unlikely h' is given h .¹³ Thirdly, $bc_D(h, e)$ gives a simple and intuitively satisfying explanation of why evidence used to calculate parameters in a theory fails to provide supporting evidence to the same extent that directly predicted evidence does (Howson, 1990). There is of course more to be said, as the extensive literature on confirmation measures testifies, and which the interested reader is invited to consult.

What about inf_R ? The confirmation function bc_R based on it is easily seen to be equal to $\log[p(e|h)/p(e)]$ where $p(e)$ and $p(h)$ are both positive, which means that where two hypotheses h and h' both entail e , bc_R depends only on $p(e)$ (it is equal to $-\log p(e)$) and so h and h' both receive the same bc_R value).¹⁴ It follows that bc_R can't in principle

¹³ See also Crupi and Tentori 2010.

¹⁴ Note that this feature is inherited also by instantiating $inf(h) - inf(h|e)$ with Bar Hillel and Carnap's measures $cont(h)$ and $cont(h|e)$ respectively where $cont(h|e)$ is defined by (1): the result is easily seen to be equal to $cont(evh)$ and thence, when h entails e , to $inf(e)$. It might seem surprising therefore, given their endorsement of (1), that a principal objective also of Bar Hillel and Carnap was to exhibit a close connection between confirmation and semantic information. The mystery is dispelled once we recall that by 'degree of confirmation' Bar Hillel and Carnap meant nothing more than conditional probability where the

discriminate between 'All emeralds are grue' and 'All emeralds are green' on a sample of green emeralds, surely a disastrous feature of any theory of confirmation seeking to model informal inductive reasoning.¹⁵ By contrast $bc_D(h,e)$ depends on the prior $p(h)$, so it naturally copes with the grue problem by allowing the prior probability of the 'grue' hypothesis to be very small, for most people probably zero¹⁶, compared with that of the other. It might be objected that 'All emeralds are grue' is equally supported by the sample, the difference between the two hypotheses being not one of empirical support but simply of priors. There seems to be a simple answer to the objection, which is that no-one should regard a contradiction as supported by any evidence at all, and presumably the same should go for any statement probabilistically indistinguishable from it.

Another well-known measure of confirmation besides bc_D and bc_R is the log-odds measure $\log[\text{odds}(h|e)/\text{odds}(h)] = \log[p(e|h)/p(e|\sim h)]$, attributed by I.J. Good (1950) to A.M. Turing who developed it during World War 2 (it has the attractive feature of adding over a conjunction $e_1 \& e_2 \& \dots \& e_n$ if the e_i are conditionally independent given h and given $\sim h$). Crupi and Tentori claim (2014, p.89), without argument, that the information-theoretic approach can be naturally extended to this, implying that it too is equal to $\text{inf}(h) - \text{inf}(h|e)$ for some information measure inf . However, the function $f(p)$ generating inf as a function of probability p is $-\log[\text{odds}(p)]$, which is arguably highly improper: it is *negative* for $1/2 < p < 1$, and, appending plus and minus ∞ to the real numbers (the enlarged set is called 'the extended real number system'), takes the value $-\infty$ on tautologies¹⁷. It is far from clear what it might mean to say that a sentence contains negative information. Nor can negative values be eliminated simply by a shift of scale. In addition, this function is concave over half its domain (in the interval $(1/2, 1)$), and as I

probability was what Carnap called 'logical probability'. Carnap called bc_D 'incremental confirmation', i.e. the *change* in Carnapian degree of confirmation.

¹⁵ Though it is endorsed by Milne (1996): one of Milne's key assumptions in his derivation of this confirmation measure is that that if e has the same probability given h and h' then h and h' are equally confirmed by it (desideratum 5, p.21). He remarks 'Viewed correctly, I submit, [the assumption] is ... utterly compelling' (1996, p.22). Utterly compelling it may seem, until viewed in both green and grue light.

¹⁶This of course raises the question of how $p(e|h)$ is to be understood if $p(h)$ is zero, given that the usual rule $p(e|h) = p(e \& h)/p(h)$ breaks down. Fortunately there are satisfactory axiomatisations of conditional probability in which $p(e|h)$ is meaningful when $p(h) = 0$, so long as h is not a contradiction (among others de Finetti 1974, volume 2, p.339; Coletti and Scozzafava 2002, p.76; Popper 1959, Appendix *iv; the measures satisfying Popper's axioms are known as Popper functions).

¹⁷ It is also not convex in $(1/2, 1)$: I will argue in section 6 that this is an important defect.

will show later, in section 6, this implies that it cannot be represented by any corresponding measure on consequence-classes (I will explain why this should be thought to be bad when we come to that discussion). It is certainly plausible that information should be a decreasing function of a suitable uncertainty measure which rates the more plausible above the less plausible, but the odds scale is extremely skewed, a transformation of the symmetric probability scale sending probability 1 to infinity while keeping 0 common to both. Also sacrificed by that transformation is the crucial property of additivity, and for these reasons the odds scale was superseded for scientific work by the probability scale, and information as a function of uncertainty is measured by probabilities and not by odds.¹⁸

So what is there to support Crupi and Tentori's claim that 'treating information and confirmation in a unified fashion is an intuitive and fruitful approach' (2014, p.81)? bc_D and bc_R may be representable as differences between ex ante and ex post measures of information, but bc_R seems fatally defective for the reasons described above, and if my argument above is correct then the Turing measure has no relation to information at all. As for bc_D , the popular choice among Bayesians, I will argue in what follows that its representation as a difference of quantities of information is actually spurious. In brief, my argument will be that, despite its distinguished pedigree in the theory of semantic information, *the measure $1-p(\cdot)$ should not qualify as an admissible measure of information at all*. This might seem an absurd claim given the virtual consensus to the contrary, but the symptoms of infirmity are already there, as we shall now see.

5. The infirmity of inf_D

The trouble starts to appear with the Bar Hillel-Carnap measure $cont(h|e)$ defined, we recall, by (1) as $cont(h\&e) - cont(e)$, and having the consequence, as we noted in section 3, that $cont(h|e) = cont(e \rightarrow h)$. That the definition (1) should engender any problems is

¹⁸ Good notes that Turing himself thought of the log-odds measure, scaled in what he called ban units, as analogous to that of *sound*, whose bel unit is the logarithm to base 10 of a ratio of sound-intensities (1950, p.63).

rather surprising, since all it does is express a fairly natural notion of excess information in h with respect to e ; moreover, as we also saw in section 3, and as should be expected of an adequate measure of excess information, $\text{cont}(h|e)$ is informationally independent of e in the sense of their common content having (cont) measure 0.

This pleasing appearance of harmony is deceptive, however, and you don't have to look far to see that something is actually very dissonant in $e \rightarrow h$ cast in the role of carrying the excess content in h over e . For suppose h is a general hypothesis of science and e the statement of the necessarily finite amount of evidence gathered so far and predicted by h . In comparison with h , $e \rightarrow h$ is an extremely *weak* statement, being implied by the already very weak $\sim e$.¹⁹ In other words, we must have $\text{cont}(\sim e) \geq \text{cont}(e \rightarrow h)$. But since $\text{cont}(e) + \text{cont}(e \rightarrow h) = \text{cont}(h)$ when h entails e , we infer that $\text{cont}(e) + \text{cont}(\sim e) \geq \text{cont}(h)$, which seems absurd in the type of case pointed out by Richard Jeffrey (1984), where h is a universal hypothesis of universal scope, and e records a relatively small finite number of its instances. But it is what (1) tells us we should accept. *Reductio!* (surely?)

At this point the reader acquainted with some of the stranger byways of philosophy of science may recall that Miller and Popper's notorious 'disproof' of probabilistic induction (1983) turned on the fact that $e \rightarrow h$ is negatively supported by e according to the bc_D measure if h entails e . Instead of seeing in this fact something amiss with taking $e \rightarrow h$ as representing the excess content of h over e , Miller and Popper bizarrely viewed it as a *reductio* of the claim that e inductively supports h , on the ground that truly inductive support must extend to the part of h 's informational content going beyond e . Since $\text{bc}_D(h,e) = \text{bc}_D(e \rightarrow h,e) + \text{bc}_D(e,e)$, the positive value of $\text{bc}_D(h,e)$ consequent on the fact that h entails e (assuming $0 < p(h), p(e) < 1$) can only be due to the confirmation of e by e outweighing the disconfirmation of $e \rightarrow h$, the (supposed) excess content. Hence, reasoned Miller and Popper, the confirmation of h by e must be the purely *deductive* confirmation of e by itself.

¹⁹ Bar Hillel and Carnap themselves noted the curious weakness of $e \rightarrow h$ (1953, p.151; they use i and j for h and e).

As virtually all commentators have agreed, this must be wrong. If, as Miller and Popper claim, and (1) endorses, $e \rightarrow h$ is truly the information in h going beyond e , then it is absurd to claim that $e \rightarrow h$ can be *countersupported* by e : e should be confirmationally neutral vis-à-vis $e \rightarrow h$, an informationally-independent piece of information ('supported' would of course have been just as bad). That granted, the obvious target of Miller's and Popper's *reductio* should surely have been the claim that $e \rightarrow h$ represents the excess information in h over e .²⁰ Small wonder then that Crupi and Tentori feel confident in maintaining their own construal of $\text{inf}_D(h|e)$: 'our reliance on inf_D throughout this paper does not imply in any way Popper and Miller's ... controversial assumption that $e \rightarrow h$ represents all of the content of h that goes beyond e ' (2014, p.82). And the Crupi-Tentori construal has, of course, the additional bonus of delivering the confirmation measure bc_D . But, *pace* Crupi and Tentori, it is no mere 'assumption' that $e \rightarrow h$ represents all of the information in h that goes beyond e . On the contrary, it is a *theorem*: a consequence of the equation (1) of arithmetical information-difference, together with the definition $\text{inf}_D(\cdot) = \text{cont}(\cdot) = 1 - \text{p}(\cdot)$. Moreover, (1) seems quite unimpeachable as a definition of information-added if the 'added' means 'arithmetically-added': what is less unimpeachable is that $e \rightarrow h$ should even represent the arithmetical difference in information between h and e . But that is what (1) says it does, *assuming that $I(\cdot)$ is equal to $1 - \text{p}(\cdot)$* . If the Popper-Miller example is a *reductio* of anything, then it seems that it can only be that assumption.

If the very definition of $\text{cont}(\cdot)/\text{inf}_D(\cdot)$ is fundamentally flawed, however, one would expect to see other pathological consequences of it which have nothing to do with (1). And indeed one does, without looking any further than (2) above. For it follows immediately from (2) and the fact that cont is non-negative, that

$$\text{cont}(a \& b) \leq \text{cont}(a) + \text{cont}(b) \quad (3)$$

with equality only when $\text{cont}(a \vee b) = 0$, i.e. $\text{p}(a \vee b) = 1$. Carnap and Bar Hillel list (3) among the properties of cont , failing however to note how strongly it conflicts with any

²⁰ Miller and Popper themselves argued for $e \rightarrow h$ as carrying the excess information of h over e on the ground that $e \rightarrow h$ is the weakest statement which, conjoined with e , delivers h as a consequence. Cf Hintikka (1968, p.313).

ordinary concept of information²¹ - a failure all the more remarkable in the light of the basic and well-understood fact about conjoining statements that one typically gets much more in the way of consequences than the individual conjuncts by themselves would suggest. We have just encountered a prime example in e and $e \rightarrow h$, both very weak statements that conjoin to give h (we are still assuming that h entails e), which can be any hypothesis as content-rich as you like. But since $ev(e \rightarrow h)$ is tautologous, (3) tells us that the content of h is nevertheless *exactly equal* to the sum of the contents of e and $e \rightarrow h$. An extreme case of this example is got by letting h be \perp itself. In some formalisations of classical logic \perp replaces \sim as a primitive symbol and negation is defined by $\sim a := a \rightarrow \perp$, so in this case $e \& (e \rightarrow h)$ can be rewritten as $e \& \sim e$. But intuitively the content generated by a contradiction (every sentence is a consequence of it) is to all intents and purposes infinite by comparison with those of e and $\sim e$. By contrast, of course, $\text{inf}_R(e \& \sim e)$ offers no such affront to intuition, being actually infinite while both $\text{inf}_R(e)$ and $\text{inf}_R(\sim e)$ are finite under the reasonable assumption that $p(e)$ and $p(\sim e)$ are positive.

Even without simple counterexamples from elementary logic, the history of science is an object-lesson in the fact that conjoining individually unremarkable assumptions can generate an explosion of remarkable, and sometimes revolutionary consequences. But that is just what (3) appears to deny, implying as it does that *any* conjunction will carry *at most* as much information as the sum of the amounts of information carried by the conjuncts separately. Inf, i.e. the $-\log p$ measure, does on the other hand satisfy the intuition that there will be non-extreme cases where according to any minimally-reasonable measure of information the conjunction of a pair, or in general a finite number, of statements contains more information than the sum of the information-contents of the statements separately.

That is all very well, it might be said, but what has information construed as unexpectedness necessarily to do with consequence-classes? I might not be at all surprised to be told that a particular logically very strong statement is true, and very

²¹ 'cont(i) is offered as one ... explicatum of the ordinary concept "amount of information conveyed by i"' (1953, p.149).

surprised to learn that a much weaker one is.²² But that purely subjective sense of unexpectedness is not the relevant one here: because of their dependence on probability measures all information functions have to obey the rule that if $Cn(a) \subseteq Cn(b)$, where $Cn(\cdot)$ is the consequence-class of ' \cdot ', then $\inf(a) \leq \inf(b)$ ²³. You may indeed be less surprised to be told that b is true than that a is, but that is because you are violating a basic rule of probability which states that if a is entailed by b , i.e. $Cn(a) \subseteq Cn(b)$, then $p(b) \leq p(a)$; and if you do this while believing that your probability evaluations are the basis for sound bets then, as Ramsey and de Finetti famously pointed out, you leave yourself open to a so-called Dutch Book, i.e. a finite set of bets that will leave you with a certain loss. Respecting the relation of consequence-class inclusion thus partially 'objectifies' both estimates of uncertainty and the information measures based on them.

Nevertheless, it is not obvious that sensitivity to the relation of consequence-class inclusion should necessarily extend to sensitivity to *measures* of consequences themselves, nor that such measures, if they exist, should be measures of unexpectedness. However, in the next section I will show that such measures do exist on consequence-classes, where the measure-terminology is justified by the fact that they are the restrictions of measures, in the technical sense, on all sets of sentences of an appropriate language L . I will show that these measures can be represented as decreasing functions $f(p)$ of a probability p for a wide class of functions f which includes $1-p$ and $-\log p$, and that they therefore qualify as measures of unexpectedness. However, it will also be shown that *the measure on consequence-classes determined by the function $1-p$ is uniformly blind to new consequences generated by conjunction.*

²² The following report in Aubrey's *Brief Lives* about the philosopher Thomas Hobbes is an amusing illustration:

He was 40 years old before he looked in on Geometry; which happened accidentally. Being in a Gentleman's Library, Euclid's Elements lay open, and 'twas the 47 El. libri I. He read the Proposition. By God, sayd he (he would now and then swear an emphaticall Oath by way of emphasis) this is impossible! So he reads the Demonstration of it, which referred him back to such a Proposition; which proposition he read. That referred him back to another, which he also read. Et sic deinceps that at last he was demonstratively convinced of that trueth. This made him in love with Geometry. (Aubrey 2015)

²³ This is in fact one of Carnap and Bar Hillel's basic rules for measures of semantic information (1952, p.12)

6. Consequences

It is an interesting fact that $Cn(a)$ was Bar Hillel and Carnap's initial choice of a 'presystematic' qualitative explication of the semantic information contained in a ²⁴ which, however, they discarded in favour of $*M(a)$ (see above, section 2), citing in the latter's favour its intuitive plausibility and being 'in accordance with the old philosophical principle "omnis determinatio est negatio"' (1952, p.11). But the main reason, they rather disarmingly claimed, is that it wins on simplicity (*ibid.*). They did not elaborate, but since the next stage of their procedure was to put a measure on whatever qualitative exemplar is selected, and since the $Cn(a)$ do not form an algebra (only an upper semi-lattice under the operation of intersection) whereas the $M(a)$ do (an algebra moreover isomorphic to the Lindenbaum sentence algebra of L), this may well be what they had in mind in preferring the $*M(a)$ explication to the $Cn(a)$ one. As we saw in section 2, any probability P on the $M(a)$ -algebra assigns a value to $*M(a)$, namely $1-P(M(a))$, inducing the measure $cont(a)$ on the sentences of L .

On the other hand, if the foregoing is correct, $cont.$ gives a strongly defective theory of semantic information, suggesting that the initial qualitative identification of the information in a as $*M(a)$, despite its claimed simplicity, was misguided. In what follows I will attempt to *demonstrate* that this is indeed the case. The argument will be developed with the help of some elementary measure theory, where despite the apparently unpromising formal structure of the consequence-classes, the measures will actually be defined directly on them. There is no paradox, and certainly no inconsistency, in this proposal: the $Cn(a)$ might not form an algebra, but the set of all subsets of the set of sentences of whatever language L is involved is an algebra, the *power set algebra* (assuming as before that the sentences are factored by equivalence), on which a finitely

²⁴ Carnap considered two possibilities: the class of nontautologous consequences versus the class of all consequences, but regarded a choice between them as unnecessary once the decision was taken in favour of the sets $*M(a)$ of possibilities excluded by a (see immediately below).

additive measure certainly exists; in fact there are infinitely many. Clearly, we can take the restriction of any such measure on that to be a measure on the $Cn(a)$.²⁵

Now we can investigate what constraints are imposed on information functions $f(p)$ by identifying them with finitely additive measures on the classes $Cn(a)$. One relation between consequence-classes we already know to be reflected in a constraint on $f(p)$: f must be a decreasing function of p to accommodate the fact that $Cn(a) \subseteq Cn(b) \Rightarrow \inf(a) \leq \inf(b)$ where $\inf(\cdot)$ is given by $f(p(\cdot))$. Also we should want f to be not only continuous but also sufficiently smooth, at least to the extent of possessing a first derivative everywhere in the open interval $(0,1)$. All this so far tells us little we don't already know. But there is a further fundamental fact about consequence-classes, namely that for all a, b ,

$$Cn(a) \cup Cn(b) \subseteq Cn(a \& b), \quad (4)$$

for which there exists (so far) no corresponding constraint on $f(p)$. And as I will show, the anomalous character of the function $1-p$ (i.e. \inf_D) derives from just this omission. Indeed, I will prove that the function $1-p$ is not just blind to the relationship described in (4), but *actually forces any measure Q on consequence-classes representing it to set at measure 0 all additional consequences created by conjunction*. There are a couple of key assumptions on which the proof will rest: (a) that $Q(Cn(a))$ satisfies the basic rules of a measure $\inf(a)$ of the semantic information in a , and (b) that there is a measure Q agreeing with the function $f(p(a)) = 1-p(a)$ on the consequence-classes $Cn(a)$. But in order not to disrupt the narrative I will give the very simple proof first and deal with the assumptions afterwards.

As a first step, the basic properties of measure tell us that

$$Q(Cn(a) \cup Cn(b)) = Q(Cn(a)) + Q(Cn(b)) - Q(Cn(a) \cap Cn(b)).$$

²⁵ Since the set of sentences of L is assumed to be countable (hardly a restrictive assumption) there are also countably additive measures on the full power set (since the existence of measures depends only on cardinality, this is equivalent to saying that there are countably additive measures on the power set of \mathbb{N} , the set of natural numbers). But countably additive measures even on measure-spaces of this cardinality are very restrictive: there can be no uniform distributions over countably infinite partitions, unlike the case of finitely additive measures (where every singleton, and every finite subset, receives the value 0 if the measure is finite). Kadane and O'Hagan (1995) investigate three familiar types of finitely additive uniform probability measure on \mathbb{N} .

But $Cn(a) \cap Cn(b) = Cn(avb)$, so we have

$$Q(Cn(a) \cup Cn(b)) = Q(Cn(a)) + Q(Cn(b)) - Q(Cn(avb)). \quad (5)$$

But if $f(p) = 1-p$ we also have

$$f(p(a)) + f(p(b)) - f(p(avb)) = f(p(a\&b))$$

whence, assuming that $f(p(a)) = Q(Cn(a))$, we infer that

$$Q(Cn(a)) + Q(Cn(b)) - Q(Cn(avb)) = Q(Cn(a\&b)). \quad (6)$$

And putting (5) and (6) together, we infer that $Q(Cn(a) \cup Cn(b)) = Q(Cn(a\&b))$! So Q does indeed assign measure 0 to the additional consequences generated by conjunction.

Conversely, it is easy to show that if Q is one of those measures Q which have this property then according to it, e and the Popper-Miller sentence $e \rightarrow h$ jointly exhaust the information in h when h entails e , however strong a statement h might be; yet $e \rightarrow h$ is a weaker sentence even than $\sim e$. For given that h entails e , the conjunction of $e \rightarrow h$ and e is equivalent to h , so that $Q(Cn(e) \cup Cn(e \rightarrow h)) = Q(Cn(h))$. Since the only common consequence of e and $e \rightarrow h$ is T , with $Q(\{T\}) = 0$, this implies that $Q(Cn(e)) + Q(Cn(e \rightarrow h)) = Q(Cn(h))$. In general however, for arbitrary Q , we have $Q(Cn(e)) + Q(Cn(e \rightarrow h)) \leq Q(Cn(h))$.

It is time to verify the two assumptions (a) and (b) stated above on which this result rests. (a), recall, is the assumption that $Q(Cn(a))$ satisfies the basic rules for a measure of the semantic information in a . This is very straightforward if we take the rules for a measure inf to be Carnap and Bar Hillel's two, (i) that $\text{inf}(T) = 0$, and (ii) that if a entails b then $\text{inf}(b) \leq \text{inf}(a)$ (1952, p.12²⁶): we simply set $Q(\{T\}) = 0$, and we know from the preceding section that (ii) is satisfied. There is just one thing to be said in addition, and that is that we need the inequality in (ii) to be strict in interesting cases, like the one above, which means that $Q(Cn(a))$ must be finite for sufficiently weak a (this may sound vague but it is little different in principle from the usual sort of Bayesian condition that where e is an evidence statement, $p(e)$ should be strictly between 0 and 1). We can't stipulate that Q *must* be finite on S however, because if we identify $Q(Cn(a))$ with $f(p(a))$ for a

²⁶ Carnap and Bar Hillel also added a third, that the information measure of a contingent sentence be positive, but this is now usually regarded as too restrictive.

continuous decreasing continuous function f from $[0,1]$ into the extended non-negative real line (i.e. the non-negative reals together with $+\infty$), we need to accommodate the case of $f(p) = -\log p$, whence $Q(S) = Q(\text{Cn}(\perp)) = -\log(\perp) = +\infty$. The possibly diverse class of information-functions $f(p)$ therefore means that we are talking not about a single measure Q but a class of measures depending on the types of function $f(p)$.

Now to assumption (b), that there is a measure Q agreeing with the function $f(p(a)) = 1 - p(a)$ on the consequence-classes $\text{Cn}(a)$. In fact we can show that the class of functions $f(p)$ representable as measures on consequence-classes *are exactly those which are convex in $(0,1)$* . First, suppose that Q is a measure and $Q(\text{Cn}(a)) = f(p(a))$. Since Q is a measure and $\text{Cn}(a) \cup \text{Cn}(b) \subseteq \text{Cn}(a\&b)$ we must have

$$Q(\text{Cn}(a) \cup \text{Cn}(b)) \leq Q(\text{Cn}(a\&b)). \quad (7)$$

Hence, from (5) and (7),

$$f(p(a)) + f(p(b)) - f(p(avb)) \leq f(p(a\&b)) \quad (8)$$

Now let $p(a) = x$, $p(b) = y$ and $p(a\&b) = x-h$, for some $h \geq 0$, and suppose that $x < y$. From the identity $p(a) + p(b) = p(avb) + p(a\&b)$ we also infer that $p(avb)$ is equal to $y+h$. We thus obtain from the inequality (8) another:

$$f(x) - f(x-h) \leq f(y+h) - f(y) \quad (9)$$

for all $h \geq 0$ and all x, y in the open interval $(0,1)$. Dividing by h , (9) is the condition for a differentiable function f to be convex in $(0,1)$ (where in addition $f(1) = f(p(T)) = 0$).

Included in the class of such functions f are both $f = 1-p$ and $f = -\log p$; excluded is the pseudo information function $-\log(p/(1-p))$, which is concave for values of p greater than $1/2$ and $f(1) = -\infty$. But we have (I hope) seen from the foregoing that the extremal case $1-p$ should also be excluded.

It is not difficult to prove the converse, that the requirement that f be convex in $(0,1)$ with $f(1) = 0$ is sufficient for the associated information function $f(p)$ to be identified with the restriction of a finitely additive measure Q on the power set of S to the set of consequence-classes. For $Q(\text{Cn}(a) \cup \text{Cn}(b))$ can simply be defined to be equal to $f(p(a)) + f(p(b)) - f(p(avb))$ for the finite additivity axiom for measures to be satisfied. All that

remains is to show that $Q(\text{Cn}(a) \cup \text{Cn}(b)) \leq Q(\text{Cn}(a \& b))$, i.e. $f(p(a)) + f(p(b)) - f(p(avb)) \leq f(p(a \& b))$. But we know that this follows from the convexity of f .

These results are to my knowledge new to the literature. As we have seen, they follow straightforwardly from the assumption that any information function $f(p(a))$ should reflect the properties of a measure Q on consequence-classes. That assumption can of course be questioned. As I pointed earlier, it follows the lead of Bar Hillel and Carnap who first selected a set-theoretical object to serve as the qualitative definition of information, and then put a measure on it which they identified with a decreasing function $f(p)$ of p : all that is different here is that Bar Hillel and Carnap chose the classes $*M(a)$ for the qualitative definition, with the deleterious consequence that the information function $f(p) := 1-p$ is then determined if the measure on the $*M(a)$ is normalised, while I have chosen their own first choice, the classes $\text{Cn}(a)$. Despite their failure to pursue it further, the case for that choice is, I believe, independently compelling: surely a fundamental feature of information, of central importance methodologically when we consider new, unifying and extending theories in science, is that in general new information is generated by the logical operation of conjunction, should be recognised by any adequate measure of semantic information. Hence the identification of $f(p(a))$ with $Q(\text{Cn}(a))$, whose power to generate novel and intuitively satisfying results has I believe amply proved itself. But the proof the pudding is not just in the eating: it also depends on the recipe. In this case both together, I believe, generate a reasonably satisfying dish.

7. Conclusion

According to Crupi and Tentori,

[Bar Hillel and Carnap's] classical analyses [of semantic information] have not remained unchallenged ... but stand as a sound basis at least for our purposes.

(2014, p.81)

In the light of the foregoing this claim must I think be strongly qualified. Of Bar Hillel and Carnap's two classic measures cont and inf of semantic information, only inf , i.e. –

logp, seems to possess the credentials of a legitimate measure, but if my earlier argument is correct then the support measure based on it is fatally defective. On the positive side, we have another way of explicating semantic information, which Bar Hillel and Carnap themselves thought the most natural before prematurely abandoning it, which, unlike inf_D , is not in conflict with the unexpectedness criterion.

References.

- Aubrey, J. 2015, *Brief Lives: with an Apparatus for the Lives of our English Mathematical Writers*, ed. K. Bennett, Oxford: Oxford University Press.
- Bar Hillel, Y. and Carnap, R. 1953, 'Semantic Information', *British Journal for the Philosophy of Science*, **4**, 147-157.
- Broad, C.D. 1926, *The philosophy of Francis Bacon: An address delivered at Cambridge on the occasion of the Bacon tercentenary, 5 October, 1926*, Cambridge: Cambridge University Press.
- Carnap, R. 1950, *Logical Foundations of Probability*, Chicago: University of Chicago Press.
- Carnap, R. and Bar Hillel, Y. 1952, 'An outline of a theory of semantic information', Technical Report no. 247, MIT.
- Coletti G. and Scozzafava R. 2002, *Probabilistic Logic in a Coherent Setting*, Dordrecht: Kluwer.
- Crupi, V. and Tentori, K. 2010, 'Irrelevant conjunction: statement and solution of a new paradox', *Philosophy of Science*, **77**, 1-13.
- Crupi, V. and Tentori, K. 2014, 'State of the field: Measuring information and confirmation', *Studies in History and Philosophy of Science*, **47**, 81-90.
- De Finetti, B. 1974, *Theory of Probability*, vol. 2, New York: Wiley.
- Good, I.J. 1950, *Probability and the Weighing of Evidence*, London: Griffin.
- Hintikka, J. 1968, 'The Varieties of Information and Scientific Explanation', *Proceedings of the Third International Congress for Logic, Methodology and Philosophy of Science*, eds. van Rootselaar and Staal, Amsterdam: North Holland.

- Hintikka, J. 1970, 'Surface Information and Depth Information', in *Information and Inference*, eds. J. Hintikka and P. Suppes, Dordrecht: Reidel, 263-387.
- Howson, C. 1990, 'Fitting your theory to the facts: probably not such a bad thing after all', *Discovery, Justification and Evolution of Scientific Theories*, ed. C. Wade Savage, *Minnesota Studies in the Philosophy of Science*, Minneapolis: University of Minnesota Press, 224-245.
- Howson, C. 2013, 'Hume's Theorem', *Studies in History and Philosophy of Science*, **44**, 339-346.
- Jeffrey, R.C. 1984, 'The impossibility of inductive probability', *Nature*, **310**, 433.
- Kadane, J.B. and O'Hagan, A. 1995, 'Using finitely additive probability: uniform distributions on the natural numbers', *Journal of the American Statistical Association*, **90**, 626-631.
- Miller, D. and Popper, K.R. 1983, 'A proof of the impossibility of inductive probability', *Nature*, **302**, 687-688.
- Milne, P. 1996, ' $\log[P(h|eb)/P(h|b)]$ is the one true measure of confirmation', *Philosophy of Science*, **63**, 21-26.
- Milne, P. 2014, 'Information, Confirmation and Conditionals', *Journal of Applied Logic*, **12**, 252-262.
- Popper, K.R. 1959, *The Logic of Scientific Discovery*, London: Routledge.