

## Hempel's Paradox, Law-likeness and Causal Relations

Severin Schroeder, *University of Reading*

It is natural to think that:

(N) A generalisation is confirmed by any of its instances.<sup>1</sup>

This has been called *Nicod's Condition*.

Thus, the general hypothesis that:

(1) All ravens are black.

would be confirmed by the observation of a black raven, reported by a statement like:

(a1) This raven is black.

However, the standard formalisation of (1) is:

(1L)  $\forall x [Rx \rightarrow Bx]$ .

Which is logically equivalent to its contra-positive:

(2L)  $\forall x [\neg Bx \rightarrow \neg Rx]$ .

In English:

(2) All non-black things are non-ravens.

This, according to (N), would be confirmed by the observation of a yellow submarine reported by:

(a2) This non-black thing is not a raven.

But it stands to reason that if (1) and (2) are logically equivalent, whatever confirms one also confirms the other. Hence, it would

---

1. I follow customary philosophical usage in using the word "confirm" in an unnaturally weak sense, where it does not mean more than: "suggest" or "provide *some* evidence for." – By an "instance" of a generalisation "All Fs are G" I mean the observation of an F that is G.

appear, paradoxically, that the observation of a yellow submarine confirms (1) the hypothesis that all ravens are black (Hempel 1965, 15). This is *Hempel's Paradox*, also called the *Ravens Paradox*.

To recapitulate the assumptions from which that implausible conclusion has been reached:

(N) [*Nicod's Condition:*]

A generalisation is confirmed by any of its instances.

(E) [*The Equivalence Condition:*]

If two hypotheses are logically equivalent, whatever confirms one confirms the other.

(C) [*The law of contra-position:*]

"All Fs are G" is logically equivalent to "All non-Gs are non-F."

Thus, our options are either to reject one or more of these assumptions, or to argue that the conclusion is not as implausible as it seems.

## I

Carl Hempel himself adopted the latter line, biting the bullet, and so did many others who attempted to give a mathematical account of evidence and probability, along Bayesian lines.<sup>2</sup> The general idea is that the observation of a yellow submarine, say, does indeed confirm the hypothesis that all ravens are black, but given the likely numbers of black and non-black things, and of ravens and non-ravens, respectively, in the domain, the degree of confirmation is much smaller than that provided by the observation of a black raven. And that, it is said, accounts for our intuitive reluctance to accept the observation of a yellow submarine as confirmation at all. As an example of this approach let us consider John Mackie's proposal.<sup>3</sup>

Mackie begins by putting forward a central tenet of Bayesian confirmation theory, which he calls the *Inverse Principle*:<sup>4</sup>

---

2. E.g.: Alexander (1958), Mackie (1963), Swinburne (1973, ch. 10), Hesse (1974, 155–162), Gaifman (1979), Chihara (1981), Horwich (1982, 54–63), Earman (1992, 69–73), Howson and Urbach (1993, 126–130); Maher (1999) and Fitelson (2006).

3. The following sections of this paper will, I hope, make it clear that my rejection of a probabilistic approach is not based on any specific defects in Mackie's proposal that more recent Bayesian attempts may have managed to avoid.

4. Cf. Horwich (1982, 57).

A hypothesis  $h$  is confirmed by an observation-report  $b$  in relation to background knowledge  $k$  if and only if the observation-report is made more probable by the adding of the hypothesis to the background knowledge. Mackie (1963, 167).

In short:  $p(b, k.h) > p(b, k)$ . Now let us say that the ratio of ravens to non-ravens is:  $x/1 - x$ ; the ratio of black things to non-black things is:  $y/1 - y$ ; and  $x < y < 1/2$ . In other words, there are fewer black things than non-black things, and fewer ravens than black things. Then, Mackie argues, we can specify the probabilities of possible observations first in relation to our background knowledge [ $p(b, k)$ ], and then in relation to the background knowledge and the hypothesis that all ravens are black [ $p(b, k.h)$ ] as follows:

$b$ :	black raven	non-black raven	black non-raven	non-black non-raven
$p(b, k)$	$xy$	$x(1 - y)$	$y(1 - x)$	$(1 - x)(1 - y)$
$p(b, k.h)$	$x$	$0$	$y - x$	$1 - y$

And we'll find, Mackie claims: that the hypothesis that all ravens are black does indeed raise the probability of observing a non-black non-raven, although much less than it raises the probability of observing a black raven.

For example, suppose (somewhat unrealistically) that the world contains 1 per cent ravens and 5 per cent black objects. Then:

$$\begin{aligned}
 x &= 0.01 \\
 y &= 0.05 \\
 1 - x &= 0.99 \\
 1 - y &= 0.95
 \end{aligned}$$

$b$ :	black raven	non-black raven	black non-raven	non-black non-raven
$p(b, k)$	0.0005	0.0095	0.0495	0.9405
$p(b, k.h)$	0.01	0	0.04	0.95

In this case, the hypothesis raises the probability of spotting a black raven from  $0.01 \cdot 0.05 = 0.0005$  to 0.01; that is by 20 times. While the probability of spotting a non-black non-raven rises from  $0.99 \cdot 0.95 = 0.9405$  to 0.95; which is (proportionally) a much smaller increase (by a factor of 1.01).

The explanation of the paradoxical appearance of the case is that “we mistake a relatively poor confirmation for no confirmation at all” (Mackie 1963, 175).<sup>5</sup>

However, it should be noted that a variant of Hempel’s paradox can be produced by another logical equivalence. (1L) “ $\forall x [Rx \rightarrow Bx]$ ” can also be transformed into:

$$(3L) \forall x [\neg Rx \vee Bx].^6$$

That is to say, (1) “All ravens are black” would be equivalent to:

(3) Everything is either not a raven or black.

But this one should expect to be confirmed by any observation of a non-raven (or anything black). So the observation of a black non-raven should confirm (3), and hence also its equivalent, our original hypothesis (1). But according to Mackie’s formulae, such an observation is made *less* probable by the adding of the hypothesis to the background knowledge.<sup>7</sup> In our example, there is a *decrease* of 0.0095 (from  $0.05 \cdot 0.99 = 0.0495$  to  $0.05 - 0.01 = 0.04$ ). So, contrary to the assumptions Mackie is trying to defend, on his account the observation of a black dinner jacket *disconfirms* the ornithological hypothesis that all ravens are black! Far from dissolving the paradox, Mackie has produced a new one.

The crucial idea behind Mackie’s approach is that there must be a certain number of black objects in the universe, even if we do not know what it is.<sup>8</sup> Call this number  $n$ , then observing that my dinner jacket is black leaves only  $n-1$  instances of black for the raven population. So it appears, marginally, to diminish the ravens’ chances of all being black. To put it in terms of the Inverse Principle, given the hypothesis (that all ravens are black), we should be more inclined to expect things of other kinds not to be black. Therefore, the observation that an object of another kind is not black confirms the hypothesis (and the observation that an object of another kind *is* black *disconfirms* it). This is what Mackie tried to employ in arguing

5. Cf. e.g. Howson and Urbach (1993, 126–130).

6. Or, if one prefers to stick to the form of a generalised conditional:

$$\forall x [[Rx \vee \neg Rx] \rightarrow [\neg Rx \vee Bx]].$$

7. Cf. Swinburne (1973, 158), Horwich (1982, 58); Maher (1999, 61).

8. This, of course, raises the problem of how one is to count unspecified objects, which, for argument’s sake, however, I shall set aside.

that Hempel's paradox is not really a paradox: only a surprising discovery. But then, by the same kind of reasoning, he is also committed to saying: that the observation of a black dinner jacket speaks *against* the hypothesis (that all ravens are black) because – roughly speaking – the more black is used up by non-ravens, the less black is left over for ravens: the less likely it is for the hypothesis (that all ravens are black) to be true. And that is not only counter-intuitive, it is also in conflict with Hempel's reasoning (which Mackie was trying to vindicate). For Hempel's reasoning leads to the opposite conclusion: that the observation of a black dinner jacket *does* confirm the hypothesis (that all ravens are black).

There are certainly circumstances when Mackie's approach is reasonable. For example, it is easy to imagine a situation when the hypothesis that all men at a party had a piece of apple cake is strengthened whenever I find a woman who had none. If, on the other hand, at that party I speak to a few women and find that none of them has a headache, I will not take that as evidence that all the men do. Headaches are quite unlike pieces of apple cake. There is not a limited number of them to go round: from which each one must be taken, leaving one less for the rest of the community. Rather, statistically speaking (i.e. leaving aside causal hypotheses), my having or not having a headache has absolutely no impact on your chances of having one.

Mackie's model for the ravens' case is something like this: There are two bags, one containing the objects of the world (ravens, dinner jackets, tulips, etc.), and one containing colour tokens. We are told in advance that only a small percentage of colour tokens are black, and an even smaller percentage of objects are ravens. And now every observation is a double draw: we take one object out of each bag. E.g.: raven : black, dinner jacket : black, &c. On this model, the colour token for my dinner jacket diminishes the store of black tokens, and hence reduces the chances for all ravens to get black colour tokens. But of course it is not like that. Colours are more like headaches than like pieces of apple cake. A more suitable model is this: Instead of the bag of colour tokens, we have a colour die. So, when we draw a dinner jacket out of the bag, we throw a die to determine its colour. If we throw "black," that does in no way diminish the chances of other objects getting "black." The throws of dice are inexhaustible, and a throw has no impact on the probabilities of the next throw (to think otherwise is called the Monte Carlo Fallacy).

By switching models in this way: from drawing things out of a bag to throwing dice, we can break the spell of the argument from the limited overall number of black objects. Similarly, for any series of 100 throws there will be a number of times a six comes up, call that number  $n$ . Again, one might be inclined to argue that every throw of a six takes away from  $n$ , and thus diminishes the chances of a six at the next throw. But here the fallacy should be obvious. Even an unbroken series of 20 sixes will not change the fact that there is a  $1/6$  chance of a six for the next throw – as for any throw of a true die.

Back to Hempel's reasoning. In general, I think it is implausible to accept the paradoxical conclusion of his argument, because that would stretch the concepts of evidence and confirmation beyond recognition. I take it to be implicit in these concepts that a given hypothesis divides possible observations into three classes: those that confirm the hypothesis (or provide evidence *for* it), those that disconfirm it (or provide evidence *against* it) and those that are irrelevant.<sup>9</sup> But according to Hempel's line of reasoning, *there could be no irrelevant observations* because

$$(1L) \forall x [Rx \rightarrow Bx]$$

is equivalent to:

$$(3L) \forall x [\neg Rx \vee Bx].$$

And everything that does not contradict (1L), is obviously an instance of (3L). In other words, everything that doesn't *refute* a hypothesis would have to be taken to confirm it (cf. Black 1966). Confirmation would boil down to logical compatibility; – but it seems clear that it does not; it is (or should be) a far more demanding concept. So the attempt to deny that there *is* a paradox I shall set aside as unconvincing.

## II

Then at least one of the three assumptions (N), (E) and (C) must be given up.

---

9. Indeed, Hempel himself sees himself as trying to clarify what evidence is *relevant* to a given hypothesis (Hempel 1965, 5).

I believe that there is a strong case against Nicod's Condition:<sup>10</sup>

(N) A generalisation is confirmed by any of its instances.

Is it really true that the observation of a yellow submarine confirms the hypothesis that:

(2) All non-black things are non-ravens.

– ? Consider a few examples of the form “This *F* is *G*.” Always assuming that it is the first *F* we have encountered, would we regard the observation as a confirmation, however slight, of the corresponding general statement: “All *F*s are *G*”?

(a4) This lady from Scunthorpe likes Jazz.

(a5) This French postman is an alcoholic.

(a6) This deep sea diver was born on 15th May.

No such observation would be taken, on its own, to confirm the corresponding generalisation. Why not?

Nelson Goodman suggests that that is because the corresponding hypotheses would not be *law-like*, but only of accidental generality.

Only a statement that is *lawlike* – regardless of its truth or falsity or its scientific importance – is capable of receiving confirmation from an instance of it; accidental statements are not. (Goodman 1983, 73)

Thus, the basic problem of confirmation theory is how to distinguish between law-like and accidental hypotheses (77). And Goodman's answer is, roughly, that for a hypothesis to be law-like it must be framed in terms of *projectible* predicates, which he then explains as those that are comparatively well *entrenched* in our inductive practice, or at least not less well entrenched than any competitors (97).

He illustrates that by an example that has become famous as *Goodman's Paradox* (or *The New Riddle of Induction*): The observation of green and only green emeralds seems not only to confirm the hypothesis:

---

10. It is actually easy to find trivial refutations of Nicod's Condition. When I already know a hypothesis to be false (or false if instantiated), it can obviously no longer be confirmed by any of its instances (cf. Good 1968). However, that does not help in the case of Hempel's Paradox where no such disqualifying background knowledge seems available.

(H1) All emeralds are green.

– but also:

(H2) All emeralds are grue.

Where the predicate “grue” is defined as: *either examined before now and green or else blue* (74). Thus, (H2) implies – very implausibly – that future emeralds will be blue! This is clearly not an acceptable case of confirmation, and Goodman’s diagnosis is that (H2) is not a law-like statement because the predicate “grue” (unlike “green”) is not properly entrenched.

This account, it seems to me, contains a lot of truth, but is not entirely right. The distinction between law-like and not law-like, or accidental, generalisations is indeed a crucial one. But contrary to what Goodman says, there is no problem about distinguishing the one from the other. Usually, the wording indicates the type of generalisation, and if it does not, we only need to consider the context or ask the speaker whether they *intended* merely to state an accidental fact or whether they meant to put forward something like a law. Consider:

- (4) Now all the children are asleep.
- (5) All children of alcoholics become neurotic.
- (6) Everything in the fridge has a bad smell.

(4) is clearly the expression of an accidental generality: there is no insinuation of any underlying law that would have licensed the prediction that all children are asleep at this point in time. (5), on the other hand, is obviously not just an observation of what happens to be the case. The idea is that there is a timeless causal regularity, like: being the child of an alcoholic makes you neurotic. Typically, law-like statements claim a (more or less) timeless, and counterfactual, validity. Accidental generalisations do not: their claim is limited to what happens to be the case in a given situation or at a given time. (6) can be meant and understood in either way. The context would be likely to disambiguate it: if the speaker continued with a detailed account of the contents of the fridge, specifying *different* reasons why different things had a bad smell, (6) would be recognisable as a merely accidental generalisation. But if instead the speaker added: “That’s why I keep my milk on the balcony,” we understand that something like a law-like claim (with counterfactual implications) was intended: “Whatever is put into that fridge takes on a bad smell.”



Now, if it is only accidental that all objects of a given set  $F$  have the property  $G$ , then there is no room for induction. Learning that one of the  $F$ s is  $G$ , or even that most of them are, will not make it any more probable that the remaining ones are. For example, if nine of a series of 10 throws of a true die are sixes, there will still be only a  $\frac{1}{6}$  chance that the tenth throw will be a six. Of course, the general statement:

(7) All 10 throws of this series will be sixes.

– becomes less unlikely with every six you throw. Initially, its probability is only  $(\frac{1}{6})^{10}$ ; after nine sixes it has increased to  $1/6$ . But that is not a matter of inductive confirmation. As Goodman plausibly says: “Confirmation of a hypothesis occurs only when an instance imparts to the hypothesis some credibility that is conveyed to other instances” (Goodman 1983, 69). Induction allows us to argue that the evidence makes it likely that *the next*  $F$  will be  $G$ : more likely than it would have been without the evidence, and if the evidence is strong: more likely than not, or even practically certain. None of this can be said in the case of a purely accidental generalisation like (7). However many positive instances we have found, that will never have any impact on the probability of the next successful throw. And whereas inductive confirmation can accumulate to the point where the probability of the general hypothesis is well above  $\frac{1}{2}$  – where the hypothesis is practically certain – in the case of purely accidental generalities, nothing short of checking each single member of the set can make it reasonable to accept the hypothesis (unless of course the probability of each single positive instance is extremely high from the very beginning).

So (serious) *inductive confirmation is only concerned with law-like statements.*

Now Goodman suggests that our observation of green emeralds does not confirm the hypothesis:

(H2) All emeralds are grue.

– because it is not a law-like hypothesis. But that is not quite right: There is no limit to the causal regularities one might (perhaps foolishly) *suspect* to hold in the world, and express by a law-like hypothesis – however bizarre or unlikely. (H2) is clearly not a *plausible* law-like hypothesis, but it is a law-like hypothesis if anybody puts it forward as such. So the problem is not that of identifying law-like

hypotheses, rather it is to identify *prima facie acceptable* law-like hypotheses, which under the circumstances merit serious consideration (even if in the end they may turn out to be false).

So the question is: Why is (H2) not a *prima facie acceptable* law-like hypothesis? – For more than one reason, I think. But the one flaw in (H2) I want to expound now is that it is *absurd*.

### III

As argued, for a single instance to confirm an empirical generalisation – to make it reasonable at least to *suspect* that such a generalisation might be true – the generalisation must not be a merely accidental one. It must be a law-like statement: so (unless it is not susceptible of any further explanation, because it is at the most fundamental level of physical phenomena<sup>11</sup>) it must be based on the assumption of *some causal regularity*.

Furthermore, questions of confirmation are never set and answered in a vacuum but in the context of at least our ordinary background knowledge. We know a fair amount about the causal regularities in the world, not only physical, but also psychological and social. So the question is not simply whether *this* observation lends credence to *that* hypothesis, but whether it does so given our background knowledge. Obviously, this rules out generalisations directly contradicted by previous observations. For instance, we cannot take the observation of a black raven as a confirmation of:

(8) All black things are ravens.

– because we have already seen plenty of black things that were not ravens (Goodman 1983, 70). But more interestingly, we will also refuse to regard a generalisation as confirmed by an observation if the generalisation involves *a type of causal regularity* we have not come across before and we could not readily explain in terms of the kinds of causal regularity we know of. For example,

(5A) All deep sea divers are born on 15th May.

---

11. E.g. that an electron has a spin of  $\frac{1}{2}$ . Such basic cases I shall set aside for the moment. They will need a different treatment. – I also set aside non-empirical generalisations, such as mathematical conjectures.

– taken as a law-like hypothesis, would imply that there is some natural or social mechanism that prevents people not born on 15th May from becoming deep sea divers. As there is no known correlation of dates of birth and physical aptitude, and no professional training in our society is made conditional on that sort of criterion, the hypothesis is so overwhelmingly improbable that a single instance counts for nothing.<sup>12</sup>

Compare this with *Goodman's view*. He would say that the predicate “born on 15th May” was not sufficiently entrenched in our

12. It might be different if we found that a curiously high number of deep-sea divers had been born on that day, but here we are only concerned with the confirmation that can be achieved by a single instance. – However, this may seem paradoxical: how can a number of positive instances provide significant confirmation for a hypothesis, while a single positive instance affords virtually no confirmation at all?

Consider two dicing scenarios:

- (A) We can be certain that all our dice are true dice.
- (B) We have reason to suspect that some of our dice may have been tampered with so that they would only ever throw the same number.

Now consider the hypothesis:

- (i) The next 100 throws with this (as yet untried) die will all be sixes.

Depending on our background knowledge, we will regard this hypothesis differently. Given scenario (A), (i) is only a possible, but extremely unlikely accidental generality. Throwing one six will, it is true, increase its likelihood a tiny bit, but the increase is quite insignificant. Even throwing 10 or 20 sixes will not make the hypothesis considerably more likely. Its chances of being true will still be nothing more than  $1/6^{80}$ .

Given scenario (B), however, positive instances will be taken much more seriously. For now we regard (i) as a plausible law-like hypothesis. It amounts to the suspicion that the die in question is made such that the six will always come up. Given that this is a possibility realistically to be reckoned with, a series of 10 sixes will make it quite rational to believe that (i) is true, and indeed even a single six, to begin with, may give rise to that suspicion.

And now consider a third scenario (C): we do not know anything about the die in question. In that case we will start off with the assumption that our die is a true die, regarding (i) as only an accidental generality and hence as overwhelmingly improbable. But when after the first 10 or 20 throws we have seen nothing but sixes, we will – even without any previous warning – have reason to suspect that this is not a true die: that, in other words, (i) may be a law-like generality based on an asymmetry in the dice. The uniformity of the results makes it rational to suspect a causal uniformity instead of the randomness of a true dice, which is a priori extremely unlikely to produce such a uniformity.

Similarly, in the case of the deep-sea divers: The hypothesis (5a) is not made significantly more likely by the observation of one instance, but it is to a serious degree confirmed by the observation of, say, 20 instances (in the absence of counterexamples). I.e., there is a leap in the confirmatory weight of the evidence: 20 instances do not just count 20 times as much as a single instance – which would still be quite insignificant. That is because the uncommon uniformity of the observed instances suggests a causal explanation that changes the status of the hypothesis: from a merely accidental generality

inductive practice to be projectible. Obviously, the overall thrust is similar: he too makes confirmation dependent on our past inductive experiences. But it seems to me misleading to see it as a matter of the characteristics of *predicates*, rather than of *types of general statements*, or causal claims.<sup>13</sup> It is not that any of the *predicates* used in my examples (a4) to (a6) are unsuitable for inductive reasoning. It is only when those predicates are used in certain kinds of general statements that the result is obviously implausible.<sup>14</sup>

As is Goodman's own example:

(H2) All emeralds are grue.

It is a combination of two highly unlikely ideas. First, it implies that our observation has a causal impact on the colour of a gemstone.<sup>15</sup>

ality to a plausible law-like hypothesis. And in the latter case, where we have some understanding of likely generative mechanisms, a positive instance has considerably greater evidential value.

13. Of course the feature of *predicates* Goodman has in mind, their use in induction, is a reflection of our causal knowledge; but then it seems preferable to go straight for that causal knowledge, rather than one of its effects.

At one point, when considering newly introduced predicates, Goodman admits as much himself: "Not predicates but certain projected hypotheses are being eliminated" (1983, 97). But strangely enough, he keeps talking as if it were a matter of adjudicating predicates.

14. You can easily imagine them figuring in some reasonable inductive arguments. For example,

The present President of the Astrological Society was born on 15th May.

So perhaps:

All Presidents of the Astrological Society were born on 15th May.

(I.e., perhaps it is required by the statutes).

(And for some of the others it is still easier to find some perfectly regular inductive employment).

Consider another case: A red-haired friend who drives an old Ford gave me a book for my birthday. Unfortunately, a week later, it was gone: I could not find it anywhere. Does this confirm the generalisation:

(F) Birthday gifts from red-haired people who drive an old Ford always vanish within a week.

– ? Obviously not. But no anomalous predicate is to blame. It is simply that the hypothesis would require some causal regularities that are in stark conflict with our background knowledge. The idea that gifts from a certain type of person always vanish within a certain space of time is absurd enough. How on earth could that be brought about? But to add that it is a certain colour of hair and the use of a certain make of car that endows a person with such magical causal powers (or is perhaps another effect of the same cause) transfers the case completely into fairyland.

15. Note that it is *not* enough to suggest that emeralds may change their colour when *exposed to light* (Clark 2002, 67), which sounds almost realistically possible as we

Secondly, the absurdity is compounded by the idea that that strange effect our eyes have on some objects comes abruptly to an end. Not, as one might expect, individually, at a certain age, when one's eyes show signs of deterioration – no: for the whole population on earth (whether infant or octogenarian) the magic stops on exactly the same day.

So, to summarise my *solution to Goodman's paradox*: When taken as a law-like statement

(H2) All emeralds are grue.

implies the assumption of some causal regularities that is patently *absurd* in the light of what we generally know and understand about causal processes. (And note that this objection is not limited to cases involving oddly contrived predicates like “grue”).<sup>16</sup>

This result amounts to a *rejection of Nicod's Condition (N)*:<sup>17</sup>

*A generalisation is not confirmed by any of its instances if it is not law-like, but only an accidental generalization. And even law-like empirical generalizations are not so confirmed if they are absurd in the light of our causal background knowledge.*<sup>18</sup>

#### IV

Back to the ravens. To dissolve Hempel's paradox, we want to show that the observation of a non-black non-raven (for instance a yellow submarine) does *not* confirm the generalisation:

are well familiar with colour changes produced by light. No, the definition of “grue” presents the much more outlandish idea that our *looking* at an object might change its colour.

It is reminiscent of some ancient ideas of vision, e.g. in Galenic physiology, according to which the eye, like a radar, sends out signals and receives their reflections by objects.

16. Contrary to Goodman's own account that holds the *predicate* “grue” responsible for the problem; likewise Quine (1969).

17. Sainsbury proposes that (N) should be replaced by:

“G4. A hypothesis ‘All Fs are Gs’ is confirmed by a body of data containing its instances, and containing no counterinstances, if and only if the data do not say that *there is*, or even that *there is quite likely to be*, a property, H, such that the examined Fs are G only in virtue of being H” (1995, 88).

But that won't do. Not even (H1) would be acceptable, as the evidence we have, suggests that emeralds are green in virtue of *some* property H.

18. Or possibly in the light of our knowledge of some basic non-causal natural laws.

(2) All non-black things are non-ravens.

(And thus, *a fortiori* not “All ravens are black”). Again, we must ask: given our background knowledge of causal regularities, would the observation of a yellow submarine make it at all reasonable to suspect, or be inclined to believe, that (2) is correct? No, we cannot accept it as confirmed by the observation of a yellow submarine. Because: for a generalisation to be confirmed by any one of its instances it must be law-like: that means (at least in the case of fairly complex phenomena, on the macro level), it must be the expression of some assumed underlying causal regularity. Hence, we need to ask: what is supposed to be the underlying causal regularity? And only if the assumption of such a causal regularity is not utterly *absurd* could a possible instance of *that* causal regularity be accepted to confirm the generalisation. Let me first illustrate that by another example:

(9) All tennis players have wrist problems.

What is the assumed causal regularity on the basis of which this claim might be supposed to be law-like? Obviously:

(i) Playing tennis causes wrist damage.

As a very familiar type of a causal process (wear and tear), this fits in well with our background knowledge. Therefore, as an expression of (i), (9) is a plausible law-like hypothesis, and an instance of (9) – being a likely instance of (i) – can be taken to confirm it. But now consider the contra-positive of (9):

(10) People without wrist problems do not play tennis.

If this is to be taken as a law-like generalisation in its own right, the intended causal claim (supposed to be instantiated by an instance of (10)) would appear to be:

(ii) Robust wrists cause people not to play tennis.

But that is highly unlikely. Why should robust wrists *prevent* people from playing tennis? We would rather expect the opposite. Hence, on this construal, (10) is not plausible as a law-like generalisation. An instance of (10) – somebody who neither plays tennis nor suffers

from wrist problems – would *not* be taken as a likely instance of the corresponding causal claim. Therefore, an instance of (10) cannot be taken to confirm it.<sup>19</sup>

Alternatively, (10) can be regarded as merely a logical variant of (9): an indirect expression of the familiar causal claim (i). But on that construal, somebody who neither plays tennis nor suffers from wrist problems would not even count as an *instance* of the hypothesis in question, i.e. (9). In other words, if the only plausible law-like construal of (9) or (10) is as an expression of the causal hypothesis:

(i) Playing tennis causes wrist damage.

– then what is supposed to confirm (9) or (10) must confirm (i): And as we are only concerned with confirmation by instantiation, it must be possible to regard it as an *instance* of (i). But obviously, a case of someone who neither plays tennis nor has any wrist problems is *not* an instance of (i); and hence not an instance, or confirmation, of either (9) or (10). The crucial point is that on the level of *causal* laws there is no law of contraposition: “A causes B” is not equivalent to “Non-B causes non-A.”<sup>20</sup>

So, it is not always true that a hypothesis “All *F*s are *G*” is confirmed by any of its instances, that is: anything that is both *F* and

19. Instead of (ii), one might perhaps suggest that (10) be construed as an expression of:

(iii) Something that causes robust wrists prevents people from playing tennis.

But that implausibly involves the idea of a constant and specific physiological cause for something we know to be heavily dependent on social factors.

20. One more example:

(11) All pillar boxes are red.

(12) All non-red objects are not pillar boxes.

Again, if these are meant to be law-like generalisations, we need to spell out the underlying causal claims. Discarding the absurd idea of a direct causal link between the properties mentioned, it is a choice between:

(i) What is to be put up as a pillar box, the authorities paint red.

and (ii) What is not red, the authorities will not put up as a pillar box.

Obviously, (ii) is absurd. You do not get post officials inspecting objects of various colours and turning their nose up at what is not red (“We can’t have *that* as a pillar box!”). Rather, whatever they make a pillar box, they give orders to paint red. So, roughly speaking, the causal direction is not from being non-red to not becoming a pillar box, but from being produced as a pillar box to being painted red. Both (11) and (12), if law-like, can only be taken as an expression of (i). Hence, they are both confirmed by instances of (i): red pillar-boxes, but not by instances of (12): such as green leaves.

G. To say which ones are and which ones are not we have to go beyond purely syntactical characterisations. We have to ask whether the generalisation under discussion is law-like: whether it is meant to express some causal regularity, and if so: which one. And this underlying causal claim is what really matters:

*An instance of a (non-basic) empirical generalization confirms the generalization only if it is also a possible instance of the underlying causal claim the generalization is meant to express.*<sup>21</sup>

Finally, in the case of the ravens, (1) and (2), we must likewise ask which of the two is a fairly direct expression of the assumed causal link, and which one is only a logical derivation, expressing the same causal claim indirectly. In other words, is it as (1) suggests: that what makes something a raven makes it black? Or is it rather that green or red (or any other non-black) colour *prevents* things from becoming ravens? Clearly the former. (The idea that *first* there is a colour, or the cause of a certain colour, which *then* determines whether or not something becomes a raven or a polar bear is absurd.) Consequently, it takes a black raven to instantiate the causal claim that *underlies* both (1) and (2); and therefore it takes a black raven to confirm both (1) and (2). A yellow submarine is not good enough.

## V

Following Goodman, I have maintained that inductive confirmation must not only raise the probability of the hypothesis in question, but make it more probable that the next instance fulfils the hypothesis, which means that the hypothesis must be a plausible law-like statement. At that point, the following objection may be raised: Even if an instance of a merely accidental generalisation (“All Fs are G”) will not increase the probability that the *next* F will be G, does it not still increase the probability of, and hence, in some sense, confirm, the

---

21. The crucial idea that in order to resolve Hempel’s Paradox we need to consider which law-like causal claim the generalisations in question can be taken to reflect and that the only plausible kind of causal claim may fail to be confirmed by instances of the contra-positive was already put forward by Rom Harré in (1970, 119–122). Hence, the general idea underlying my solution to Hempel’s Paradox is not completely new, but whereas Harré argues that attention to “the generative mechanisms at work” undermines the Equivalence Condition, I construe it as a reason to reject Nicod’s Condition.



generalisation? And is this minimal sense of “confirm” not sufficient to engender Hempel’s Paradox? Could it not still be maintained, paradoxically, that the observation of a yellow submarine will give us a little more reason to believe that all ravens are black?

The answer to the last question is clearly no: An increase in probability does not always give us more *reason to believe* something. For example, an observation that, by removing one potential counterexample, increases the probability of a hypothesis from  $1/999,999$  to  $1/999,998$  does not thereby give a rational person any more reason to take the hypothesis seriously, let alone to believe it to be true.

We should not lose sight of the fact that Hempel’s Paradox occurs in a *scientific* context. The question is whether the observation of a single positive instance can give us a reason to believe that a certain hypothesis might well be true: that at least it merits to be considered as a serious contender for truth. But this attitude is incompatible with regarding the hypothesis as the expression of a merely accidental generality (e.g. that all in a long series of throws of true dice are sixes). For calling something merely “accidental” *means* that there is no good reason to expect it to occur; that the truth of such a generalisation would be a mere fluke. As far as the scientific consideration of a hypothesis is concerned, once it has been classified as merely an accidental generalisation “all bets are off”: it is no longer in the running; and increases in probability become entirely insignificant. So if the term “confirmation” is not to be uprooted entirely from its scientific habitat; from the question of rational credibility, it must mean considerably more than an increase in probability.

Even so, it would be paradoxical enough if Hempel’s Paradox could be put in terms of probability alone. Could one not still insist that the observation of a yellow submarine made it slightly more probable that all ravens are black? The explanation would be that what appears to be an irrelevant observation is not after all *entirely* irrelevant since every observation of an object that is not a non-black raven reduces the number of possible counterexamples; and the lower the number of possible counterexamples the higher the probability of the hypothesis.

The first weakness in this line of reasoning is that it presupposes that the number of objects in the universe is finite, which may be doubted. But let that pass. The real flaw in the argument is this: It presupposes that, initially, we regard every (non-black) object we are going to see as a potential counterexample to the hypothesis: a

potential non-black raven; and therefore, the more non-black non-ravens we see, the likelier it becomes that all non-black objects are non-ravens (i.e., that there are no counterexamples to the original hypothesis (1)). But that presupposition is false. *The world is not a bag out of which we draw one object after another in such a way that any object is equally likely to be drawn at any time.* We already know a lot about the kinds of objects the world contains and their possible or likely location. I know, for example, that there are no ravens in a cash machine. Hence, the fact that any non-black things coming out of it are not ravens is simply part of my background knowledge and so cannot in the slightest increase the probability of our hypothesis. In fact, in most circumstances of my everyday life I *know beforehand* that the next object I am going to see is not a raven; therefore no such observation can plausibly be taken to diminish the class of possible counterexamples to hypothesis (1).

Once again, it might be helpful to replace the misleading model that engendered the paradox by a more appropriate one. Our everyday observations of objects in the world are not like observations of objects drawn randomly out of a bag; they could more plausibly be likened to drawing objects out of various *different* bags, of which, for the most part, we already know what their likely contents are. One of them contains birds, but most of them contain quite different kinds of things, such as banknotes, carpets or blades of grass. Evidently, a draw from a bag known to contain no birds (e.g. my wardrobe) can neither instantiate our hypothesis (1), nor can it provide a counterexample to it. But if it is known from the outset that such a draw cannot provide a counterexample to the hypothesis, then its not doing so cannot strengthen the hypothesis either (it cannot be regarded as indirectly relevant as an elimination of a potential counterexample). In short, the contents of a bag known to contain no birds have no bearing on the hypothesis that all ravens are black.

Proponents of a Bayesian approach are not likely to be impressed by this objection. Indeed some of them are well aware that “observing (known) non-ravens cannot tell us anything about the color of ravens” (Fitelson 2006, 97). They just insist that instead of observing objects we already know to be non-ravens, we need to sample objects at random from the universe. Then, according to the standard Bayesian line, it can be shown that randomly sampling a non-black non-raven slightly increases the probability that all ravens are black.

That may be so, but even apart from the insignificance of the minute probabilities involved, the idea of such a cosmic lottery (where the draw of a raven is a priori as possible as the draw of a planet) is so utterly artificial and unfeasible that it is hard to see what bearing the results of this set-up could have on our (scientific or everyday) concerns with questions of evidence and probability.<sup>22</sup>

### References

- Alexander, H. G. (1958). "The Paradoxes of Confirmation." *British Journal for the Philosophy of Science* 9: 227–233.
- Black, M. (1966). "Notes on the 'Paradoxes of Confirmation'." In J. Hintikka and P. Suppes (eds.), *Aspects of Inductive Logic*. Amsterdam: North-Holland, pp. 175–197.
- Chihara, C. (1981). "Quine and the Confirmational Paradoxes." In P. A. French, T. E. Uehling Jr. and H. K. Wettstein (eds.), *Midwest Studies in Philosophy. Vol. 6: The Foundations of Analytic Philosophy*. Minneapolis: University of Minnesota Press, pp. 425–452.
- Clark, M. (2002). *Paradoxes from A to Z*. London: Routledge.
- Earman, J. (1992). *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. Cambridge, MA: MIT Press.
- Fitelson, B. (2006). "The Paradox of Confirmation." *Philosophy Compass* 1: 95–113.
- Gaifman, H. (1979). "Subjective Probability, Natural Predicates and Hempel's Ravens." *Erkenntnis* 14: 105–147.
- Good, I. J. (1968). "The White Shoe *qua* Herring is Pink." *British Journal for the Philosophy of Science* 19: 156–157.
- Goodman, N. (1983). *Fact, Fiction and Forecast*, 4th edition. Cambridge, MA: Harvard.
- Harré, R. (1970). *The Principles of Scientific Thinking*. London: Macmillan.
- Hempel, C. G. (1965). *Aspects of Scientific Explanation*. New York: The Free Press.
- Hesse, M. (1974). *The Structure of Scientific Inference*. London: Macmillan.

---

22. I am indebted to David Dolby, John Preston and Hugh Rice for their helpful comments on earlier drafts of this paper.

- Horwich, P. (1982). *Probability and Evidence*. Cambridge: Cambridge University Press.
- Howson, C. and P. Urbach (1993). *Scientific Reasoning: The Bayesian Approach*, 2nd edition. Chicago, IL: Open Court.
- Mackie, J. (1963). "The Paradox of Confirmation." *British Journal for the Philosophy of Science* 13: 265–277. repr. in P. H. Nidditch (ed.), *The Philosophy of Science*, Oxford: OUP, 1968.
- Maher, P. (1999). "Inductive Logic and the Ravens Paradox." *Philosophy of Science* 66: 50–70.
- Quine, W. V. O. (1969). "Natural Kinds", in his *Ontological Relativity and Other Essays*. New York: Columbia University Press.
- Sainsbury, M. (1995). *Paradoxes*, 2nd edition. Cambridge: Cambridge University Press.
- Swinburne, R. G. (1973). *An Introduction to Confirmation Theory*. London: Methuen.

*Department of Philosophy*  
*University of Reading*  
*Whiteknights*  
*Reading RG6 6AA*  
*s.j.schroeder@reading.ac.uk*