CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Ghent University Academic Bibliography

BLACK RAVENS, WHITE SHOES AND SCIENTIFIC EVIDENCE. THE RAVENS PARADOX AND/IN SCIENTIFIC PRACTICE

Erik Weber, Mathieu Beirlaen & Inge De Bal

Centre for Logic and Philosophy of Science
Ghent University (UGent)
Blandijnberg 2, B-9000 Gent , Belgium

Abstract
A well-known consequence of Hempel's account of confirmation is the Ravens Paradox. In this paper we discuss this paradox from the viewpoint of scientific practice. The main worry, when looking at this paradox from a scientific practice perspective, is that it seems to lead to problematic methodological advice for scientists: it seems to licence 'indoor ornithology'. We show that this problematic advice only follows from Hempel's account if one adopts a suboptimal view of what counts as evidence for an hypothesis. We present and defend a more sophisticated view of what counts as evidence, which takes random sampling – an important methodological principle in scientific practice – into account. On this sophisticated view, the problematic methodological advice connected to the Ravens Paradox is avoided

Keywords
Carl Hempel; confirmation; indoor ornithology; Ravens Paradox; positive evidence.

BLACK RAVENS, WHITE SHOES AND SCIENTIFIC EVIDENCE.
THE RAVENS PARADOX AND/IN SCIENTIFIC PRACTICE

1. Introduction

The Ravens Paradox was put forward by Carl Hempel in an article in *Mind* (1945) which was reprinted with a postscript in his book *Aspects of Scientific Explanation and other Essays in the Philosophy of Science* (1965). The paradox has been discussed by many philosophers since then. A clear and concise formulation can be found in a recent book on epistemic inference:

> Suppose one seeks to confirm the hypothesis, *H*, that all ravens are black. On the positive instance account, one at least initially looks around for ravens; they constitute appropriate instances of the hypothesis and are positive if also black. But *H* is also logically equivalent to *H'*, that all non-black things are non-ravens. So anything which is neither black nor a raven is a positive instance of, hence confirms *H'*. If it is required, as seems entirely reasonable, that objects confirming a hypothesis confirm all hypotheses logically equivalent to it—the "equivalence condition"—there emerges the paradoxical result that whatever is not black and not a raven, from white shoes to red herrings, also confirms the hypothesis that all ravens are black. (Bandyopadhyay et al. 2016, p. 127)

Hempel's theory of confirmation is a positive instance account, and thus leads to this paradoxical result.

In the last decades, there has been an intense debate among Bayesian epistemologists about what is the best way to deal with the Ravens Paradox within a Bayesian framework. This debate has led to increasingly sophisticated Bayesian solutions of the Ravens Paradox. This paper does *not* fit into this tradition.

The main issue discussed in this paper different. We are primarily interested in the implications (if any) of the Ravens Paradox for scientific practice. The following quote by Nelson Goodman points out the problem range we are interested in:

> The prospect of being able to investigate ornithological theories without going out in the rain is so attractive that we know there must be a catch in it. (1983, p. 70)

Indeed, the Ravens Paradox seems to imply that, if you want to gather evidence for the hypothesis 'All ravens are black', it suffices to look around in your room and come up with a correct observation report describing some non-black non-ravens you find in your surroundings. Goodman calls this 'indoor ornithology' (1983, p. 71). The question is: would such indoor ornithology be good scientific practice? Goodman suggests a negative answer: he thinks there must be a 'catch'.

What is at stake is the validity of the following methodological advice we could (as philosophers of science) give to scientists:

> *Indoor Ornithology Advice*
> When considering the hypothesis that all ravens are black, regard white shoes, as well as black ravens, as positive evidence.

The main aim of this paper is to argue that, despite the Ravens Paradox, there is no reason at all why we should give this advice to scientists. In order to reach this aim we proceed as follows. In Section 2 we summarise Hempel's account of confirmation and explain in detail how it leads to the Ravens Paradox. In Section 3 we argue for the importance of random sampling in the context of hypotheses such as 'All ravens are black'. In Section 4 we show that a corollary of taking random sampling seriously is that indoor ornithology is banned.

In Section 5 we point out that our line of reasoning in Section 3 and 4 challenges a hidden assumption in many discussions of the Ravens Paradox: the identification of positive evidence with confirmation. We show that by distinguishing the two, the Ravens Paradox is defused: it becomes an *unproblematic* result, rather than a paradoxical result. In this way we reach the second aim of this paper, which is to show that the so-called Ravens Paradox is not a paradox at all. This second aim is a derivative one: it is about a very important corollary of the way we tackle our main aim.

2. Hempel's account of confirmation and its implications

2.1 Hempel defines confirmation as a relation between an observation report and an hypothesis. His view can be summarised in two definitions (cf. Hempel 1965 p. 37):

(C1) Observation report O *directly* confirms hypothesis H if and only if O entails the development of H for the class of those objects that are mentioned in O.

(C2) Observation report O confirms hypothesis H if and only if H is entailed by a class of sentences each of which is directly confirmed by O.

To illustrate the first definition, consider the following observation report (where R stands for 'raven' and B for 'black':

(I) $Ra \wedge Ba$
$Rb \wedge Bb$
$Rc \wedge Bc$

In this report, three objects are mentioned: *a*, *b* and *c*. In general the *development* of an hypothesis for a finite class of objects is defined as that what this hypothesis would assert if only the objects in that class existed (1965, p. 36). We apply this definition to the hypothesis 'All ravens are black', which we can write formally as:

(II) $(\forall x)(Rx \rightarrow Bx)$

The development of this hypothesis for the class {a, b, c} is:

(III) $(Ra \rightarrow Ba) \wedge (Rb \rightarrow Bb) \wedge (Rc \rightarrow Bc)$

Observation report (I) entails statement (III), which is the development of hypothesis (II). So, according to the Hempelian definition (C1), report (I) directly confirms hypothesis (II).

2.2 In order to explain how Hempel's account leads to the Ravens Paradox, we have to bring in non-black non-ravens. For sake of clarity, it is useful choose to focus on a subclass of those. In our exposition we use white shoes. However, nothing depends on this choice: other non-black non-ravens would be equally good to develop the argument.

Hempel's account implies the following:

> *White Shoe Confirmation*
> The hypothesis 'All ravens are black' is confirmed by an observation of a white shoe.

There are two ways in which Hempel's account leads to *White Shoe Confirmation*. The first is indirect and makes use of Hempel's equivalence condition, which states that every observation report that confirms one of two logically equivalent sentences, also confirms the other hypothesis (1965, p. 13). This property of Hempel's account follows straightforwardly from definition (C2). Since 'All ravens are black' is logically equivalent to 'All non-black objects are non-ravens', every observation report that confirms (II) will also confirm:

(IV) $(\forall x)(\neg Bx \rightarrow \neg Rx)$

The converse is also the case: every observation report that confirms (IV) also confirms (II). Now consider object *d* which is a white shoe. For this object we can observe the following:

(V) $\neg Bd \wedge \neg Rd$

This report directly confirms (IV) because it entails its development for {d}, which is:

(VI) $\neg Bd \rightarrow \neg Rd$

If we apply definition (C2), the result we get is that observation report (V) not only confirms hypothesis (IV) – this is regarded by Hempel as unproblematic – but also hypothesis (II). So we have *White Shoe Confirmation*: an observation report about a white shoe confirms the hypothesis 'All ravens are black'.

There is also a more direct way in which Hempel's account leads to *White Shoe Confirmation*, a way that involves only (C1): according to this definition, observation report (V) directly confirms hypothesis (II) because $Rd \rightarrow Bd$ (= the development of (II) for {d}) follows logically from (V).

As already noted, nothing in the argument nothing depends on the objects being white shoes in particular. All non-black non-ravens will do. Hempel's account also implies, among others that, the hypothesis 'All ravens are black' is confirmed by an observation of a blue chair or a red sock.

3. What counts as positive evidence for conditional hypotheses?

3.1 With 'conditional hypotheses' we mean statements of the form $(\forall x)$ $(Px \rightarrow Qx)$, for instance 'All ravens are black' and 'All swans are white'. We present a theory about what counts as evidence for such hypotheses. So from now on, in this paper, every hypothesis H is assumed to have this form (both in the definitions we give and in our examples).

Suppose that, with respect to such conditional hypotheses, we formulate the following advice to scientists:

> *Hempelian Generic Advice*
> When considering a conditional hypothesis H, regard an observation report O as positive evidence for H if and only if it confirms this hypothesis (in the sense defined by (C1) and (C2)).

The reason why we call this a 'Hempelian' advice is clear: it uses his definitions of confirmation. We call it 'generic' because it is about all conditional hypotheses.

The upshot of Section 2 is that the *Indoor Ornithology Advice* (which is not a generic advice, but a specific one about one hypothesis) follows inevitably from this generic advice. If we want to avoid endorsing indoor ornithology, we have to come up with an alternative generic advice. This is what will do in 3.2. In 3.3 we defend this alternative generic advice. In Section 4 we will show how it avoids the *Indoor Ornithology Advice*.

3.2 Like the Hempelian advice, our alternative advice uses two definitions. The first definition is that of *direct positive evidence*. It goes as follows:

(PE1) Observation report O constitutes direct positive evidence for hypothesis H if and only if
(a) O is obtained as the result of a random sampling procedure devised for testing H; and
(b) O entails the development of H for the class of those objects that are mentioned in O.

According to this definition, whether O constitutes direct positive evidence for H depends on how O was obtained and on the logical relation between O and H. We also use a definition of indirect positive evidence:

(PE2) Observation report O constitutes *indirect* positive evidence for hypothesis H if and only if O constitutes direct positive evidence for hypothesis H' from which H is deductively derivable.

With these definitions in place, we can formulate our alternative:

> *Sophisticated Generic Advice*
> When considering a conditional hypothesis H, regard an observation report O as positive evidence for H if and only if it constitutes direct or indirect evidence for this hypothesis (in the sense defined by (PE1) and (PE2)).

We call this advice 'sophisticated' because we think it is better than the Hempelian alternative. We argue for that in Section 3.3.

3.3 Random sampling is standard practice when scientists try to estimate population parameters, such as the relative frequency of a property Q in a population P (this is what is done in a conditional hypothesis of the form $(\forall x)$ $(Px \rightarrow Qx)$: the hypothesis is that the relative frequency equals 1). Here is a quote from a methodological textbook on social research which illustrates this:

> Probability sampling is where every individual element in a population is chosen at random and has a known, non-zero chance of selection. […]
> Probability methods of sample selection are the best if the researcher wishes to describe accurately the characteristics of a sample in order to estimate population parameters, for example to establish the needs of older people or the attitudes of residents in an area to a re-development scheme. (Arber 2001, p. 61)

The message in the second quote (from a different methodological textbook) is the same:

> Probability sampling relies on the use of random sampling. […] The theory behind its use is that the best way to get a *representative sample* is to ensure that the researcher has absolutely no influence on the selection of people/items to be included in the sample. The sample, instead, should be based on completely random selection from the population being studied. (Denscombe 2010, pp. 24-25; italics in original)

In order to illustrate why random sampling is important, imagine that I go to the zoo in my hometown and observe that the cage with ibis birds contains 100 red ibises. At least three scenarios are possible:
(a) I have reasons to think the sample is the result of a random selection procedure applied of the population of all ibises. Therefore, I consider the observation of 100 red ibises as positive evidence for the claim 'All ibises are red'.
(b) I have no idea about how the sample came about. Nevertheless, I consider the observation of 100 red ibises as positive evidence for the claims 'All ibises are red'.
(c) I know that the zoo management has selected red ibises (instead of white, green and/or black ones) for aesthetic and/or ecological reasons. Nevertheless, I consider the observation of 100 red ibises as positive evidence for the claims 'All ibises are red'.
Scenario (a) would be good scientific practice, while (b) and (c) would be bad scientific practice. In scenario (b) we infer a population property from a sample while we have no idea at all whether the properties of the sample allow such inductive inferences. That is bad scientific practice. Scenario (c) is even worse: we infer a population property from a sample while we know that the sample is *not* suited for such inferences. That is also bad scientific practice.

We can now show why our alternative advice is better than the Hempelian proposal. According to the *Hempelian Generic Advice* the three scenarios (a) till (c) are all fine. Our *Sophisticated Generic Advice* allows only scenario (a) and puts a ban on (b) and (c). In other words: our advice leads to good scientific practice, while the Hempelian advice allows good was well as bad scientific practice.

4. Implications for indoor ornithology

Let us stake stock. We have argued that the Hempelian methodological advice, which would allow indoor ornithology, is a bad advice, for reasons independent from the Ravens Paradox. We have proposed a better advice. In this section we show that a corollary of this alternative advice is that indoor ornithology is banned.

4.1 Suppose we are seeking evidence for the claim ‘All ravens are black’. Given that (∀x) (Rx → Bx) and (∀x) (¬Bx → ¬Rx) are logically equivalent, there are – in principle – two ways to proceed:
(1) Compile a representative sample of ravens. Check how many are black. If all are black, conclude that (∀x) (Rx → Bx).
(2) Compile a representative sample of non-black objects. Check how many are non-ravens. If all are non-ravens, conclude that (∀x) (¬Bx → ¬Rx). Derive (∀x) (Rx → Bx) from this by logical deduction.
The first route constitutes a direct strategy: we try to obtain an observation report that constitutes *direct* evidence (as defined in (PE1) for the target hypothesis. The second route is an indirect strategy: we seek direct positive evidence for an hypothesis that is logically equivalent to the one we are interested in (cf. definition (PE2)).

The first route is not easy but feasible (ravens live in a wide array of environments; for other species of birds the analogous exercise is much easier). The second route is a non-starter because there is no clear demarcation of what an “object” is. Suppose someone asks you to collect a sample of 1000 objects that is representative of all non-black objects in the universe. This seems an impossible task, given that objects are everywhere (all over the Earth but also elsewhere in the Universe) and exist at all levels: living beings, organs, cells, molecules, …; libraries, books, pages in books, …; computers, screens of computers; liquid crystals in the screen, … .

The unfeasibility of the second route has important implications. In order to clarify this, we spell out the consequences – for the hypotheses ‘All ravens are black’ and the logically equivalent ‘All non-black objects are non-ravens’ – of this impossibility, in combination with our definitions (PE1) and (PE2):
(1) Black ravens provide direct positive evidence for ‘All ravens are black’ if and only if they are in an observation report that is the result of a random sampling procedure devised to test this hypothesis.
(2) Because ‘All non-black objects are non-ravens’ follows logically from ‘All ravens are black’, black ravens provide indirect positive evidence for ‘All non-black objects are non-ravens’ if and only if the random sampling condition in (1) is satisfied.
(3) Non-black non-ravens never provide direct positive evidence for ‘All non-black objects are non-ravens’ because of the impossibility to construct a random sampling procedure for testing this hypothesis.
(4) Non-black non-ravens never provide indirect positive evidence for ‘All ravens are black’ because they fail to provide direct evidence for ‘All non-black objects are non-ravens’ (cf. (3)).
The consequences (3) and (4) tell us that white shoes, red pencils and other things we may find by looking around in room are worthless evidence: they have no evidential value for ‘All ravens are black’ nor for ‘All non-black objects are non-ravens’. Hence, indoor ornithology is banned.

The asymmetry in feasibility of random sampling routes is a generic phenomenon. The consequences spelled out above remain valid – mutatis mutandis – if we substitute alternative hypotheses for the hypotheses ‘All ravens are black’ and ‘All non-black objects are non-ravens’. For instance, it is easily checked that they remain in force if we substitute ‘white’ for ‘black’ and ‘swans’ for ‘ravens’.

4.2 In his book *Theory and Reality* Peter Godfrey-Smith has pointed out that random sampling can shed light on the Ravens Paradox. He writes:

> If the white shoe was encountered as part of a random sample of nonblack things, then it *is* evidence” (2003, p. 215)

Our argument in Section 4.1 implies that white shoes can never be part of a random sample of non-black things. What Godfrey-Smith says refers to a situation that never occurs. Because of the difference in feasibility we have pointed at in in 4.1, the use of random sampling procedures excludes indoor ornithology. From our perspective, Godfrey-Smith emphasises something important (viz. that the procedures by which we arrive at observation reports are relevant) but misses an important point. When considering what we have call ‘the second route’ in 4.1, he characterises it as a “more unusual approach” (2003, p. 215). In our view, it is an impossible approach. This latter insight is crucial, because it is an indispensable ingredient of our argument that against *Indoor Ornithology Advice*.

In the next section we will explain how our findings defuse the Ravens Paradox. Before we tackle this, it is useful to point at an asymmetry. Refutation of universally quantified conditional hypotheses does *not* require random sampling. A white raven counts as evidence *against* the hypothesis that all ravens are black, no matter how we arrived at this observation. In other words, it is perfectly possible to define falsification of hypotheses of the form “All P are Q” without a random sampling clause:

(NE) Observation report O constitutes direct negative evidence for hypothesis H if and only if O directly falsifies hypothesis H.

Godfrey-Smith also saw this asymmetry. He writes:

> A black raven refutes the hypothesis that no ravens are black, regardless of the procedure behind the observation. (2003, p. 2016)

5. Summary

On Hempel's theory of confirmation, an observation report consisting of white shoes, blue chairs, and other non-black non-ravens confirms the hypothesis that all ravens are black. We argued that, despite initial appearance, this is unproblematic for the practicing scientist: it does not license indoor ornithology. Our argument rests on the inclusion of an additional demand: make sure that observation reports are obtained via the right kind of procedure for gathering evidence, i.e. a procedure that involves random sampling. We have defended the legitimacy of this demand in Section 3. In Section 4 we have explicated in detail how, once this demand is met, indoor ornithology is banned.

References

Arber Sara (2001), 'Designing Samples', in Nigel Gilbert (ed.), *Researching Social Life* (2nd edition). London: Sage Publications, pp. 58-82.

Bandyopadhyay Prasanta, Brittan Jr. Gordon & Taper Mark (2016), *Belief, Evidence, and Uncertainty. Problems of Epistemic Inference*. Springer

Denscombe Martyn (2010), *The Good Research Guide* (4th edition). Maidenhead: McGraw Hill Open University Press.

Godfrey-Smith Peter (2003), *Theory and Reality*, Chicago: University of Chicago Press.

Goodman Nelson (1973), *Fact, Fiction and Forecast (3rd edition).* Indianapolis - New York: The Bobbs-Merrill Company.

Hempel Carl (1945), 'Studies in the Logic of Confirmation', *Mind* **54,** pp. 1–26 and 97–121.

Hempel Carl (1965), *Aspects of Scientific Explanation and other Essays in the Philosophy of Science*. New York: Free Press.