The Ravens Revisited

PETER LIPTON

> And here it is constantly supposed, that there is a connection between the present fact and that which is inferred from it. Were there nothing to bind them together, the inference would be entirely precarious.
>
> *David Hume*[1]

Astronomers study the behaviour of the stars; philosophers of science study the behaviour of the astronomers. Philosophers of science, alongside historians and sociologists of science, are in the business of accounting for how science works and what it achieves. There is more to the philosophy of science than principled descriptions of scientific activity, since there are also all the normative questions of justification and warrant, but the descriptive task is an important part of the discipline and the primary focus of the present essay.

Science is one of the most complex and impressive things our species does, so it is natural to wish to understand how it works. Worthwhile but relatively easy, one might think. For unlike stars, astronomers can talk. To find out how science works, it might seem that all that is required is to get a few cooperative scientists to describe what they do. But things are not so easy, in part because there is such a large gap between what people can do and what they can describe. It is one thing to be able to ride a bicycle; it something quite different to be good at describing the physics and the physiology behind that ability. The same contrast between doing and describing applies to more sedentary pursuits. It is one thing to be expert at distinguishing grammatical from ungrammatical strings of words in one's native tongue; it is something quite different to be able to specify the principles by which these discriminations are made. The same applies to science. It is one thing to be a good scientist; it is something quite different to be good at giving a general description of what scientists do.

[1] D. Hume, *An Enquiry Concerning Human Understanding*, T. Beauchamp (ed.) (Oxford: Oxford University Press, 1748/1999), section IV.

https://doi.org/10.1017/S1358246100009759 Published online by Cambridge University Press

Scientists are not good at the descriptive task. This is no criticism, since their job is to do the science, not to talk about how they do it. Philosophers of science are not very good at describing science either: this is more embarrassing, since it is their job. It turns out that giving even an approximately correct general description of what are apparently the simplest aspects of scientific research and judgement is remarkably difficult. For example, philosophers of science have worked long and hard to provide a general account of the apparently simple tripartite distinction between data that would be taken to support a given hypothesis, data that would tell against it, and data that would be irrelevant; but at least most of these attempts turn out to have the highly implausible consequence that the set of data irrelevant to a given hypothesis, the set one would have thought was the largest of the three by some margin, is empty. This difficulty is brought out with particular force by the raven paradox of confirmation, a problem that will occupy us for most of this essay.

Problems of Induction

The great Humean problem of induction is in the first instance about justification rather than description.[2] Our beliefs about things in the world we have not observed are based on inconclusive reasons. Hume's description of our inferential practice is glib: according to him, we suppose that the unobserved is like the observed. His main question is not what our practice is like, but what entitlement we have for following it. Given his description of that practice, for him the question of entitlement is tantamount to the question of what reason we have to believe that the unobserved is similar to the observed, what reason we can have to believe that nature is uniform. Hume's notorious answer is that no reason for the uniformity of nature is possible, because the uniformity claim is itself about the unobserved. In particular, observation could not teach us that the unobserved is like the observed, since observation could only teach us about the unobserved if we already knew that the unobserved is like the observed. You can't get there unless you have already arrived.

[2] D. Hume, *An Enquiry Concerning Human Understanding*, T. Beauchamp (ed.) (Oxford: Oxford University Press, 1748/1999), sections IV, V.

https://doi.org/10.1017/S1358246100009759 Published online by Cambridge University Press

The problem of induction is a problem of epistemic disconnection. If Hume is right, what we see gives us no guidance about what we have not seen. The observation of a black raven gives no more warrant for saying that the next raven will be black than for saying it will be blue. Every conjunction is arbitrary. Of course that is not the way we think: we use the observed as a guide to the unobserved, and we distinguish arbitrary conjunctions from those where the first conjunct is a guide to the second. Of course the probability of an arbitrary conjunction will go up if we observe what was before an uncertain conjunct, but that is mere 'content cutting', where discovering the truth of one conjunct does not affect the probability of the other. Genuine confirmation, by contrast, occurs when observing one conjunct increases the probability of the other.[3] Hume's sceptical argument is that content cutting is all we ever get, and that we have no basis for giving any determinate probability to the unobserved. But that is not at all the way we think. As Hume emphasised, we think that observing a black raven increases the probability that the next raven will be black: we think there is such a thing as inductive confirmation. We distinguish what we regard as arbitrary conjunctions—where one conjunct tells us nothing about another—from somehow connected conjuncts, where one gives information about the other. Our inductive practices make a clear distinction between genuine confirmation and mere 'content cutting'. We start by knowing part of the content of a statement is correct; in some cases we take that to provide a reason to believe that the rest of the content is correct too; in other cases we do not. Whether or not this distinction is justifiable, whether or not there can be any justification for using the observed as a guide to the unobserved, the descriptive question remains. On what basis do we make this discrimination between genuine confirmation and mere content cutting?

It is perhaps surprising that, like the problem of justification, the project of describing our inductive practices also throws up a problem of disconnection. Here the problem is in a way the mirror image of the problem of justification, for it is the problem of coming up with an account of confirmation that is not wildly over-permissive, incorrectly counting mere content cutting as if it were genuine confirmation. On initially plausible accounts, confirmation would be much too easy, so that A would count as confirming A&B however unrelated the two conjuncts may be. The

[3] N. Goodman, *Fact, Fiction and Forecast*, 4th ed. (Indianapolis: Bobbs-Merrill, 1983), 68–69.

https://doi.org/10.1017/S1358246100009759 Published online by Cambridge University Press

challenge is to find an account that reflects the distinction we make in practice and so to avoid counting mere content cutting as confirmation, and to make sense of real confirmation.

The problem of arbitrary conjunctions arises immediately for a simple version of the hypothetico-deductive model of confirmation. On that model, the truth of the observed consequences of a hypothesis confirms the unobserved consequences. As a description this seems corroborated by a great deal of scientific practice. But it is an immediate consequence of the model that every statement counts as confirming any other, since for any conjunction A&B, A will count as confirming B.

An obvious response to this difficulty is to place a restriction on the logical form of the hypothesis to be confirmed, a restriction that rules out arbitrary conjunctions. This is what Nicod's criterion does, insisting that the hypothesis to be confirmed take the form not of a conjunction but of a universal conditional, such as 'All ravens are black' or 'All emeralds are green'.[4] But then along came Nelson Goodman, whose new riddle of induction supplied a nefarious mechanism for converting arbitrary conjunctions into the syntactic form of universal conditionals.[5] Goodman's famous example is 'All emeralds are grue', where the novel predicate refers to green things observed before a specified future time and blue things not so observed. Varying his example slightly, we may start with the arbitrary conjunction 'observed emeralds are green and unobserved ravens are black'. Green emeralds obviously do not confirm the prediction that unobserved ravens will be black, and Nicod's criterion does not immediately fall into the trap of saying they do, since this is a conjunction, not a universal conditional. But what Goodman taught us is that, by dint of modest creative predicate construction, this conjunction can be recast into the universal condition 'All emeravens are grack', and Nicod's criterion succumbs after all, mistakenly counting green emeralds as confirmation of black ravens.

The Raven Paradox

The other notorious puzzle of confirmation, the raven paradox, poses a similar challenge for the description of our inductive

[4] C. Hempel, *Aspects of Scientific Explanation* (New York: Free Press, 1965), 10–13.

[5] N. Goodman, *Fact, Fiction and Forecast*, 4th ed. (Indianapolis: Bobbs-Merrill, 1983), chapter III.

https://doi.org/10.1017/S1358246100009759 Published online by Cambridge University Press

practices.[6] The problem is generated by the irresistible pull of the equivalence condition on confirmation, according to which whatever observation confirms an hypothesis must also confirm any logically equivalent hypothesis. (Irresistible, because if we see that two statements are logically equivalent, we see that they must stand or fall together, so any evidence for one is immediately evidence for the other.) Nicod's criterion says that 'All ravens are black' is confirmed by black ravens, and that 'All non-black things are non-ravens' is confirmed by things that are neither ravens nor black. But since those two hypotheses are logically equivalent—positive and contrapositive—the observation of green emeralds and white pieces of paper now count as if they too confirmed the hypothesis that all ravens are black. Once again, something has gone badly wrong. Taken together, the new riddle induction and the raven paradox reveal a lack of discrimination in our accounts of confirmation from both sides. The new riddle shows the evidence confirming too many hypotheses; the raven paradox shows a hypothesis being confirmed by too much evidence.

A common first reaction to the raven paradox is that we have somehow failed to register the obvious fact that the hypothesis that all ravens are black is only about ravens, and so only the colouration of ravens bears on its truth value. But as logical sophisticates know, a universal conditional actually makes a claim about every object in the universe. Another logical equivalent to 'All ravens are black' is 'Everything is either black or not a raven'. If anything in the universe fails to meet this specification, the raven hypothesis is false. Thus the paradox seems to show not just that some conjunctions are not confirmed by any of their conjuncts, but that confirmable conjunctions are in a sense not confirmed by all of their conjuncts. The raven hypothesis is something like a conjunction across every object in the universe: it claims that each of them has the property of being either black or not a raven. But what is the difference between those conjuncts that confirm and those that don't? How does the green emerald differ from the black raven? Both are things that satisfy the disjunctive property, black or not a raven. In the balance of this essay I want to consider briefly three approaches to solving the raven paradox, approaches to explaining why, as Goodman put it, there is no prospect for indoor

[6] C. Hempel, *Aspects of Scientific Explanation* (New York: Free Press, 1965), 14–20.

https://doi.org/10.1017/S1358246100009759 Published online by Cambridge University Press

ornithology.[7] Why do some conjuncts confirm while others do not? The first two of the approaches I will consider are familiar, and in my view at once enlightening and problematic; the third is a sketch of one alternative, inspired by reliabilist approaches in epistemology.

Bayesians: Differential Priors

One popular solution to the raven paradox is Bayesian.[8] Bayesians hold that belief is a matter of degree and can be represented in terms of probabilities. Thus p(E) is the probability I give to the statement E, which may range from 0, if I am certain E is false, to 1, if I am certain E is true. By representing beliefs as probabilities, it is possible to use the mathematical theory of probability to give an account of the dynamics of belief, and in particular an account of inductive confirmation. The natural thought is that evidence E supports hypothesis H just in case the discovery of E causes me to raise my degree of belief in H. To put the point in terms of probabilities, E supports H just in case the probability of H after E is known is higher than the probability of H beforehand. In the jargon, what is required is that the posterior probability of H be greater than its prior probability.

What makes Bayesianism exciting is that the standard axioms of probability theory yield an equation that appears to tell us just when this condition of confirmation is satisfied, and so to give us a precise theory of induction. That equation is Bayes's theorem, which in its near simplest form looks like this:

$$p(H|E) = p(E|H)p(H)/p(E)$$

On the left hand side, we have the conditional probability of H given E. Bayesians treat this as the posterior probability of H, so the figure on the left hand side represents the degree of belief you should have after the evidence E is in. The right hand side contains three probabilities, which together determine the posterior. The first of these—p(E|H)—is the probability of E given H, known as the 'likelihood', because it represents how likely H would make E. The other two probabilities on the right hand side—p(H) and

[7] N. Goodman, *Fact, Fiction and Forecast*, 4th ed. (Indianapolis: Bobbs-Merrill, 1983), 71.

[8] Cf. C. Howson & P. Urbach, *Scientific Reasoning: The Bayesian Approach* (La Salle: Open Court, 1989).

https://doi.org/10.1017/S1358246100009759 Published online by Cambridge University Press

p(E)—are the priors of H and E respectively. They represent degree of belief in hypothesis H before the evidence described by E is in, and degree of belief in E itself before the relevant observation is made. This process of moving from prior probabilities and likelihood to posterior probability by moving from right to left in Bayes's theorem is known as conditionalising and is claimed by the Bayesian to characterise the dynamic of degrees of belief and so the structure of inference.

To get the flavour of the way Bayes's theorem works, consider a hypothetico-deductive case; that is, a case where H deductively entails E. Here the likelihood $p(H|E)$ is simply 1, since if H entails E then E must be true given H. Under this happy circumstance, Bayes's theorem becomes even simpler: the posterior of H is simply the prior of hypothesis divided by the prior of the evidence: $p(H)/p(E)$. Bayes's theorem thus tells us that a successful prediction of H confirms H, so long as the correctness of the prediction was neither certainly true or false before it was checked. For so long as the prior of E is less 1 but more than 0, $p(H)/p(E)$ must be greater than $p(H)$, and this is to say that the posterior of H will be greater than the prior of H, which is just when the Bayesian claims there is inductive confirmation.

Bayesianism seems at first to succumb to the raven paradox rather than solving it. For as the raven hypothesis entails that the next thing I see will be either a raven or not black, and this claim has a prior probability less than unity, the Bayesian is committed to the claim that whatever I see next will count as confirming the raven hypothesis so long as it does not contradict it, whether it be black raven, green emerald, or white paper. And indeed this is what the Bayesian says. But since confirmation on the Bayesian scheme is a matter of degree, it is open to the Bayesian to maintain that while most observations compatible with the raven hypothesis would only confirm it to a negligible degree, observing a black raven provides substantial confirmation.

Not only can the Bayesians this, but they have a good reason to say it. For the difference in confirmation here comes down to a difference in the priors of the evidence—the lower the prior, the greater the confirmation—and there is good reason to claim that the probability that the next raven I see will be black is substantially lower than the probability that the next non-black thing will be a non-raven. To see why this is so, notice that if there is a counterexample to the raven hypothesis, it will be a non-black raven, and so will be a member both of the class of ravens and of the class of non-black things. But given how many fewer ravens

https://doi.org/10.1017/S1358246100009759 Published online by Cambridge University Press

there are than non-black things, the probability of finding this counterexample by looking in the class of ravens is much greater than the probability of finding it by looking in the class of non-black things. The probability that a member of the class of ravens will be black will thus be lower than the probability that a member of the class of non-black things will be a non-raven, so observing a black raven confirms more than observing a non-black, non-raven.

Bayesianism is a powerful approach to describing our inductive practices, and as we see it has something enlightening to say about the raven paradox in particular. But I have three reservations about this solution. The first is that Bayesianism has been confronted with various general difficulties, not all of which it can clearly meet. To mention what is perhaps the best known of these, it faces the problem of 'old evidence'.[9] On the Bayesian scheme, what powers the increase from the prior to the posterior probability of H is a transition in the probability of the evidence from its prior value to its value after observation (effectively unity). But this contrast is only generated on the standard Bayesian scheme if H is formulated before E is observed. Otherwise, by the time H arrives on the scene its prior already reflects the impact of the earlier observation of E, so we have no before/after observation contrast. Thus without some special jiggling of the Bayesian scheme, a black raven observed before the scientist generates the raven hypothesis provides even less confirmation for that hypothesis than the subsequent observation of a piece of white paper, since it provides no confirmation at all.

A second reservation concerns the challenge with which I began this essay, the challenge of distinguishing genuine confirmation from mere content cutting. It is not clear that Bayesianism adequately illuminates that distinction. This can be seen in the context of the new riddle of induction. Recall the arbitrarily conjunctive hypothesis 'All emeravens are grack'. The Bayesian scheme entails that the observation of green emeralds does confirm this hypothesis, which seems the wrong result. In response to this, the Bayesian may say that the trouble with these gruesome hypotheses is that they have themselves such a low prior probability that even allowing that they are confirmed by their instances does

[9] Cf. J. Earman, *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory* (Cambridge, MA: MIT Press, 1992), chapter 5.

https://doi.org/10.1017/S1358246100009759 Published online by Cambridge University Press

not lead them to end up with high probability.[10] But this is doubly wrong. First of all, even in a case where the posterior probability remains low, it is still seems wrong to say that there is confirmation. Secondly, as the example of 'All emeravens are grack' shows, gruesome hypotheses may have a high prior. To alter the example, consider a hypothesis with a small number of instances, such as 'everyone in the room is left-handed'. This may have a reasonably high prior probability, if there are few enough people in the room, and the Bayesian scheme will say it is confirmed by each person in the room we then find to be left-handed, yet this is a model of mere-content cutting, since the probability that the remaining people are left-handed does not go up.

I do not say that there is no way for the Bayesian to clarify the distinction between content cutting and genuine confirmation. For example, while the Bayesian notion of hypothesis confirmation is in itself only a kind of content cutting, presumably the Bayesian can also avail herself of various conditional probabilities that will make the distinctions we want. Thus I may assign higher probability to the claim that unobserved ravens will be black if observed ravens are black than to the claim that unobserved ravens will be black if observed emeralds are green. But this takes me to my third reservation. Even if the Bayesians can find a way of describing a scientist's web of probability assignments so as to reflect the inductive distinctions she makes, it is not clear how much light is shed on the sources of these degrees of belief. Bayes's theorem may be well seen as a kind of consistency constraint, the probabilistic analogy to a requirement that we not contradict ourselves.[11] Here the prior probabilities and the likehoods are like the premises of a deductive argument, and the application of Bayes's theorem like the application of valid rules of deduction. But just as a consistency constraint does not say much about the source of the premises, so the application of Bayes's theorem does not tell us much about where the priors come from. Bayesians may be happy at this stage to maintain that subjective probabilities are just that; but it is I think worth investigating whether there are not any additional and more objective constraint on confirmation that our inductive judgements attempt to track.

[10] C. Howson, *Hume's Problem: Induction and the Justification of Belief* (Oxford: Oxford University Press, 2000), 239–240.

[11] C. Howson, *Hume's Problem: Induction and the Justification of Belief* (Oxford: Oxford University Press, 2000), chapter 7.

https://doi.org/10.1017/S1358246100009759 Published online by Cambridge University Press

Peter Lipton

Quine: Unprojectible Complements

As Hume remarked in the quotation I have used as the epigraph for this essay, our inductive practices seem based on the presumption of a connection between the instance we observed and the instance we infer. The raven paradox shows that not only is the presumption difficult to justify, it is difficult even to acknowledge. The Bayesian analysis of induction, for all its merits, does not seem to do very well on this score. So I turn now to another well-known solution that may be more promising. This is the invention of Willard van Orman Quine, presented in brief compass at the start of his essay on natural kinds.[12]

Quine proposes an assimilation of the raven paradox to the new riddle of induction. Like many, including Goodman himself, Quine takes it that the new riddle shows that only certain predicates are suitable for induction. 'Emerald', 'green', 'raven', 'black' and their ilk are 'projectible'; but 'grue', 'emeraven' and 'grack' are not. Quine's clever proposal is that the complement of a projectible predicate is not projectible: thus 'non-black' and 'non-raven' are unprojectible, just like 'grue'. So just as the observation that past emeralds have been grue gives no reason to believe that future emeralds will be grue, so the fact that observed non-black things have been non-ravens gives no reason to believe that unobserved non-black things will be non-ravens. As Quine observes, this proposal can be made compatible with the equivalence condition on confirmation, if we say that black ravens confirm 'All ravens are black' and all its logical equivalents, and that white paper confirms none of them.

Quine's solution is appealing in part because it suggests a notion of connection behind the distinction we draw between cases of genuine confirmation and arbitrary conjunctions. Ravens tell us about other ravens, because ravens form a natural kind; non-black things do not tell us about other non-black things, because non-black things form a heterogeneous class, a rabble not a kind. But of course there are difficulties with this proposal too.

The raven paradox is not a puzzle about birds. To solve it, it is not enough to find a way to discount the epistemic power of non-black, non-ravens: the solution must generalise. Quine of course recognises this, which is why he suggests that the complement of every projectible predicate might be unprojectible.

[12] W. V. Quine, 'Natural Kinds', in *Ontological Relativity* (New York: Columbia University Press, 1969), 114–116.

https://doi.org/10.1017/S1358246100009759 Published online by Cambridge University Press

But there are a number of difficulties with this suggestion. First of all, though this may not be a problem, it is worth noting that this proposal will not give a general account of unprojectibility, since many unprojectible predicates, such as 'grue', do not have projectible complements either. Not all the complements of unprojectible predicates (e.g. 'not grue') are projectible. More problematically, it is not clear that the all complements of projectible predicates are unprojectible either. Consider for example the hypothesis that all massless particles move at the speed of light. This, I take it, contains only projectible predicates. But I would say the same about its contrapositive that everything that does not move at the speed of light has mass. Or take the hypothesis that all invertebrates lack kidneys: this seems no more or less instance confirmable than its contrapositive.

A different sort of difficulty for Quine's proposal is that projectibility of predicates is not the whole story about the confirmability of hypotheses, since the same predicates may appear in both confirmable and unconfirmable hypotheses. Thus hypotheses such as 'All gold spheres have a diameter of less that one hundred miles' is a standard sort of example of a merely accidental truth not confirmed by its instances (though in this case we have excellent independent reasons for thinking it true, given our knowledge of the total amount of gold), yet there is nothing wrong with the predicates it contains to prevent them from figuring in other, instance confirmable hypotheses. Again, we may have inductive evidence for highly accidental generalisations. Thus if we sample randomly from a room of students, the observation that everyone in the sample is left handed does confirm the hypothesis that the others are left-handed too. Here what counts seems to be the method of sampling rather than the status of the predicates.

Like many philosophical problems, the raven paradox initially presents itself as a case where we know what the right answer ought to be, even if we cannot figure out how to arrive at it. Surely we know that other people do have minds, even if we do quite see how this knowledge is possible in the face of the sceptical problem of other minds. Similarly, it seems obvious that black ravens confirm the raven hypothesis and white paper does not; the trouble is in coming up with a description of our inductive practices which entails that result. But in fact it is not so obvious what the right answer is in this case. The paradox gains its initial power because we are inclined to say both that black ravens confirm 'All ravens are black' and that non-black non-ravens confirm 'All non-black things are non-ravens': we are inclined to agree with Nicod that

https://doi.org/10.1017/S1358246100009759 Published online by Cambridge University Press

hypotheses are confirmed by the observation of their instances. We are then discombobulated by the observation that these two hypotheses are logically equivalent. Thus the temptation to argue that the two hypotheses, correctly construed, are not really equivalent, so one can hold on to the judgement that each kind of evidence is confirmatory of its proper hypothesis. But the Bayesians, as we have seen, end up with a different result, according to which the right answer is that non-black, non-ravens provide only negligible confirmation for *either* hypothesis. And Quine goes even further, claiming on different grounds that they provide no confirmation at all.

So it is not so clear what the right answer is in this case, and I am not convinced that the answer Quine reaches is the one we should wish, since in fact the demarcation between confirming and non-confirming instances does not appear to follow the distinction between positive and contrapositive instances. The example of mass and speed is perhaps a case in point. Another and more pervasive illustration is given by a central feature of our inductive practices, dubbed by John Stuart Mill 'the method of difference'.[13] What Mill observed is that we very frequently make inferences from effects to causes on the basis of a *pair* of instances, one positive, one contrapositive. Thus in a controlled experiment the evidence that a certain drug lowers cholesterol is based on the combination of positive instances of people taking the drug whose cholesterol goes down with contrapositive instances of people whose cholesterol does not go down and who are not taking the drug but only a placebo. The inference depends on the appropriate pairing of instances, and has much greater force than either kind of instance on its own. In everyday life, application of something like the method of difference is ubiquitous, for example when we manipulate a cause to determine its effects. So an answer to the raven paradox that would rule out all contrapositive instances is probably not the answer we want.

My final reservation about Quine's solution parallels my concern that the Bayesian has not adequately illuminated the basis of judgements of confirmability. Quine's appeal to natural kinds seems more promising in this regard, because of the intuition that instances that are members of the same kind bear a deep similarity, and that this underpins our willingness to infer from what we observe in one member of the kind to others. But there are

[13] J. S. Mill, *A System of Logic*, 8th ed., London: Longmans, Green and Co., 1904) section III.VIII.2.

https://doi.org/10.1017/S1358246100009759 Published online by Cambridge University Press

metaphysical difficulties in saying in what this similarity might be taken to consist. Moreover, it is not even clear how some metaphysically robust notion of natural kinds would help in this context. Grant that ravens are fundamentally like each other in a way that non-ravens are not. It is not clear why exactly this should matter when it comes to confirmation, given the logical equivalence with which the raven paradox began, and given that if ravens are united in what property they have, then non-ravens must be united in what property they lack.

Reliabilism: The Tracking Condition

An alternative approach to the Raven paradox that I now sketch appeals to a reliabilist epistemology which locates warrant in reliable belief-forming mechanisms. In contrast with an internalist epistemology that would make justification necessary for knowledge, where justification is a matter of reasons and argument, reliabilism is an externalist epistemology according to which what makes some true beliefs into knowledge is that the cognitive mechanism that gave rise to those beliefs was in fact reliable, not that the knower has reasons for thinking it so. Perceptual knowledge provides a good illustration of the reliabilist perspective. I know that there is a mug on my desk, not because I have made a warranted inference from a muggish experience to the existence of a mug in the world, nor because I have some general justification for relying on the testimony of my senses, but because I believe there is a mug on my desk, there is, and the perceptual mechanism by which the belief was acquired was reliable. Not to put too fine a point on it, my eyes are at the centre of a good mug-detection system, and that is what matters.

Reliabilism raises many questions and has been much discussed. I must ignore almost all of this in the present brief advertisement, but there are two issues I need to flag. One concerns the way the notion of reliability is to be understood. What does it mean to say that a belief-forming mechanism is reliable? One might for example try to analyse this in terms of a tendency or of a long-run frequency of truth generation, but here I will instead follow Fred Dretske and Robert Nozick and understand reliability as applying

https://doi.org/10.1017/S1358246100009759 Published online by Cambridge University Press

to the single case, in terms of a counterfactual.[14] In its simplest form, the idea is that a true belief that p was reliability generated just in case had p not been true, it would not have been believed. Thus my belief that there is a mug on the table counts as reliably generated because, had there not been a mug on the table, I would not have believed that there was. To use Nozick's term, a belief that satisfies this sort of counterfactual 'tracks' the truth.

Another point about reliabilism that needs flagging is that while this form of externalism allows for knowledge without reasons or inference, it also applies to inferential knowledge. Inferential practices are themselves belief-forming mechanisms, and reliability provides a way of thinking about what makes something into a warrant-conferring inference. Reliabilism thus also provides a way of thinking about what makes something good evidence. What we want is an evidential policy that will lead us to accept hypotheses that are true, and where those true beliefs satisfy the tracking condition. We want to turn ourselves into reliable detectors of the truth value of the hypothesis in question: evidence helps us to do that.

We want a general epistemic policy that tends to take us towards truths and away from falsehoods. And I am suggesting that we also want our true beliefs to track. When we believe a true hypothesis, we also want it to be the case that we would not have believed it unless it were true. So when we believe H true because of evidence E, what we want to be the case is that, if H had not been true, we would not have inferred H from E. In one sort of simple case, this would amount to the condition that if H had not been true, E would not have been true either.[15] It is particularly easy to see how this is supposed to work in the case of an inference from symptom to underlying condition. Koplic spots are good evidence that someone has measles, because it is easy to determine that someone has koplic spots, because someone who has koplic spots is likely to have measles and, to come to the tracking condition, if that person did not have measles, she would not have had koplic spots either. And in the case where the hypothesis in question is a universal conditional, one way of construing a tracking condition is that our evidential situation be such that, had the hypothesis been false, we

[14] F. Dretske, 'Conclusive Reasons', *Australasian Journal of Philosophy* **67** (1971), 1–22; R. Nozick, *Philosophical Explanations* (Cambridge, MA: Harvard University Press, 1981), 172–196.

[15] R. Nozick, *Philosophical Explanations* (Cambridge, MA: Harvard University Press, 1981), 248.

https://doi.org/10.1017/S1358246100009759 Published online by Cambridge University Press

would have seen a counterexample. A slightly different form the condition might take that has certain advantages for our purposes is that, had there been an unobserved counterexample to the hypothesis, there also would have been an observed counterexample. Thus where our observations give us good reason to believe that all ravens are black, it should be the case that had there been unobserved non-black ravens, there also would have been observed non-black ravens.

When our belief in a hypothesis satisfies the tracking condition, we have gotten ourselves into a situation where we would not have believed that hypothesis if had been false. Our evidential practices supported this by making it the case that we have looked in a place where H would have failed, had it been false. From this perspective, good evidence practice contributes to tracking, or makes tracking more likely.[16] My suggestion is that tracking is a confirmational virtue. This is not to say that tracking is the whole story of confirmation. For one thing, the sort of condition I have indicated only applies to true hypotheses, and we also need to make sense of the confirmation of hypotheses that are false. Good empirical technique should lead us to reject hypotheses that are false as well as lead us to establish a tracking relationship with hypotheses that are true. Nor would I claim that data that do not abet tracking must therefore be confirmationally irrelevant: they may have other epistemic benefits.

Nevertheless, the tracking condition may help to illuminate the distinction between genuine confirmation and mere content cutting, because it provides a connection between the observed and the unobserved. This is illustrated by the measles example: there is a counterfactual connection between the observed symptom and the unobserved disease. We can also see this connection in the context of a general hypothesis. Suppose that we truly believe a general hypothesis on the basis of observing some of its instances. The tracking reliabilist will say that still we do not know that the hypothesis is true if, had it been false, we still would have observed only positive instances. What is required for knowledge is that if the hypothesis had been false, we would have noticed, because we would have observed a counterexample.

Let us now apply this tracking reliabilist perspective directly to the raven paradox. As we have noted, all the objects in the world are

[16] There is more work to be done to articulate how this probability is to be understood, but I would like to capture the idea that confirmation in its tracking element is a matter of degree.

https://doi.org/10.1017/S1358246100009759 Published online by Cambridge University Press

in a sense instances of the hypothesis, and will have the disjunctive property of being either black or a non-raven if the hypothesis is true. But it does not follow that every such object contributes to tracking, or contributes to tracking to the same extent. So far as the tracking condition is concerned, those instances that confirm do so to the extent that they would have been likely to be counterexamples if the hypothesis had been false. So the question to ask is this: What objects in the world would have been non-black ravens if not all ravens were black? I suggest that these objects would lie among the ravens, and not among the pieces of paper. A white piece of paper would not have been a non-black raven, had not all ravens been black, whereas a black raven might have been. Roughly speaking, the idea is that in the closest possible world where the raven hypothesis is false, actual black ravens are more likely to be non-black ravens than actual non-black non-ravens are likely to be non-black ravens.[17] And if this is right, it is not just because there are so many more ravens than there are non-black things that black ravens carry more evidential punch than white pieces of paper. Consider the hypothesis that all the ravens in a room (or in a possible world) are black, and assume that there are the same number of ravens in the room as there are non-black things in the room. Still, I would say that black ravens confirm the hypothesis and brown chairs do not. And I'm suggesting this is linked to the fact that it is the ravens rather than the non-black things that would have been non-black ravens, had there been any.

The search for evidence that would track is related to the Popperian idea that the scientist should look where a hypothesis is most likely to fail if it is false, but the conditions are not the same. For example, if we already know that some objects are black, we know they cannot be counterexamples to the raven hypothesis, so they would be of no further interest to the Popperian. But some of these black things (i.e. the ravens among them) may well contribute to tracking. So the tracking condition helps to account for the fact that it is sometimes worth gathering evidence even when it is known that the evidence will not contradict the hypothesis. Conversely, the observation of potential falsifiers, such as things antecedently known to be non-black which turn out on inspection also to be non-ravens, may not help much with tracking.

On the tracking view of confirmation, the reason that one conjunct of an arbitrary conjunction does not confirm the other is

[17] This way of putting the point was given to me by Paul Dicken.

https://doi.org/10.1017/S1358246100009759 Published online by Cambridge University Press

that there is no tracking relationship between them.[18] The fact that past emeralds are green gives no reason to believe that future ravens are black, because even if those future ravens had not been black, the past emeralds would have been green. So even positive instances of some hypotheses—such as 'All emeravens are grack'—do not confirm them. This applies also to less contrived cases. Thus the tracking condition may help to explain why repeating the same experiment over and over again eventually yields diminishing returns, confirmationally speaking. At first the experiment improves the prospects of tracking the truth of the hypothesis being tested, but repetition beyond a certain point yields diminishing returns, because the later repetitions do not make tracking more likely. By contrast, variety in evidence—a traditional evidential virtue—does improve the prospects for tracking. (Various other standard evidential virtues, such as number and precision, may also have a tracking aspect.)

The picture of the tracking condition for confirmation that I have sketched so far is primarily of some instances tracking other instances: observing black ravens confirms the raven hypothesis because if some unobserved ravens had been non-black, then some of the observed ravens would have been non-black too. But the tracking condition has a broader application. For example, it helps us to see how observations that are not intuitively part of an hypothesis may nevertheless confirm it. It is not that only similar instances track each other: inferences from effects to causes may be particularly good trackers, as the reliabilist approach to symptom and disease suggests. The tracking condition may also fit well with the idea that evidence confirms hypotheses that would explain it well.[19] A good explanation makes a difference to the phenomenon being explained, which is to say that if the explanatory facts had been otherwise, the phenomenon would not have occurred. And this is tantamount to the condition that, had the explanatory

[18] As Tim Williamson pointed out to me, this shows why it is not fully adequate to construe the tracking condition as saying simply that if the hypothesis had been false the evidence would have been different. For suppose that the hypothesis is an arbitrary conjunction of the observed and the unobserved, where the observed conjunct is more 'fragile' or contingent than the unobserved conjunct. In this case the simple tracking condition would be satisfied even though there is no confirmation. What is required I think is that the observed track the unobserved, which is not the case for this conjunction.

[19] Cf. P. Lipton, *Inference to the Best Explanation* 2nd edition (London: Routledge, 2004).

https://doi.org/10.1017/S1358246100009759 Published online by Cambridge University Press

hypothesis been false, the evidence would have been different. And the tracking condition may help to illuminate the various forms of hedged realism, where it is argued that one should not believe the full content of a high-level scientific theory. For it may be that the available evidence only helps to establish a tracking relationship with a part of the full theory: it is only if that part had been false that we would have noticed.[20]

The tracking condition is applicable beyond the context of instantial confirmation in yet further ways. For the probability of tracking depends not just on the content of the evidence, but also on facts about how it was gathered. Take the case of the over-zealous lab assistant. Knowing how ego-invested I am in my pet hypothesis that all ravens are black, my assistant adopts a policy of only showing me ravens that are black, a policy of quietly suppressing any counterexamples he should come across. As it happens, he never has to implement his strategy, since all ravens are in fact black; nevertheless, the black ravens I am shown do not provide strong evidence that all ravens are black since, in part because even if there had been counterexamples, I never would have seen them. On the positive side, the tracking condition may help to account for the fact that a policy of random sampling makes it possible to confirm certain hypotheses that are 'accidental'. Thus polls may give good evidence about the distribution of views on some topic even though the hypothesis that the views have that distribution is not particularly lawlike, because the sampling strategy satisfied the condition that if that views had been differently distributed, the poll results would have been different too.[21]

How does this tracking perspective on confirmation compare with the Bayesian and the Quinean approaches to the raven paradox? Unlike the Bayesian, the tracker appeals to a genuine connection between instances, not just content cutting. And where Bayesian probabilities are subjective, tracking relations are objective, or anyway more objective than brute degrees of belief. And where the Bayesian looks for evidence with a low prior probability, the tracker looks for what will improve the prospects for tracking, and these are different features. On the one side there can be evidence that does not have low prior but tracks, such as old evidence that tracks. On the other side there is also evidence that

[20] Cf. A. Chakravartty, 'Semirealism', *Studies in History and Philosophy of Science* **29** (1998), 391–408.

[21] Cf. J. Woodward, 'Realism about Laws', *Erkenntnis* **36** (1992), 200.

https://doi.org/10.1017/S1358246100009759 Published online by Cambridge University Press

has a low prior but does not track, such as an arbitrary and unlikely conjunct. And for the tracker, when it comes to classes, size isn't everything: members of the larger class may contribute to tracking and members of the smaller class may not.

Unlike Quine, for the tracker what is important is not the predicate per se, but whether one instance tracks another. But it may be that natural kinds are precisely classes where members do tend to track each other. And perhaps we may say that a predicate is projectible when one instance tracks another. But the tracker can also make sense of the confirmational value of contrapositive instances, for example in the context of Mill's method of difference. One of the reasons that the positive-contrapositive pairs of instances in a controlled experiment can provide such strong support for a causal hypothesis is because well-chosen pairs of this sort are very likely to track. Thus in a well-controlled experiment to test whether a drug lowers cholesterol, where the outcome is positive, we would not have gotten those results if the drug did not really have that effect. Either the cholesterol of the drug takers would not have decreased, or the cholesterol of the placebo takers would have decreased. The combination of well paired positive and contrapositive instances maximize the chances of tracking. This also helps to explain why what makes for good pairing in the method of difference depends on relevant similarity between the members. We don't want the people in the control group to have a very different diet or a very different level of physical activity than those in the group that gets the drug. If the drug takers also take unusual amounts of exercise and have an exceptionally healthy diet, we might well have gotten the contrasting experimental results even if the drug had not lowered cholesterol, and that is one reason why such an experiment would not provide good evidence.

Scientists want to get themselves into the happy cognitive situation where they know that certain hypotheses are true. To do this, they must find hypotheses that are indeed true, but that is not enough. They must also hook up with those hypotheses in a tracking relationship, and this is something that good evidence and good epistemic practices enable them to do. As I have suggested, to see how they do this, suppose that the hypothesis in question is true. Now consider how the world would have been different if the hypothesis had been false. Scientists are themselves part of the world. In order to know the hypothesis to be true, scientists need to get themselves into a situation such that in the nearby world where the hypothesis is false, it is also the case that they do not believe it to be true. Certain observations and practices help them to achieve

https://doi.org/10.1017/S1358246100009759 Published online by Cambridge University Press

this; others do not. Observing black ravens may help, if the hypothesis in question is the raven hypothesis; but this is not simply because black ravens are instances of the hypothesis. And pieces of white paper may not help, even though they are instances of the contrapositive. What matters is what would have been different and what the scientist would have noticed.

Conclusion

This has been a lightning sketch of how one might use the tracking approach to illuminate aspects of empirical confirmation in general and how it might help with the raven paradox in particular. But I do not suffer from the illusion that tracking is the whole story about confirmation, or that the account will work for all cases in the rough form here presented. There are also challenging questions about the epistemology of the counterfactuals on which the account depends. How do we judge whether a tracking relation is likely to hold? And there are other tricky features of counterfactuals that raise challenges that a properly developed tracking account would have to meet. For example, since evidence is in the past and prediction is in the future, the tracking account requires that we allow counterfactuals to backtrack, so that it may be true to say that if were to have been different in the future, things would also have been different in the past.[22] A different sort of worry arises from the possibility that some empirical hypotheses are necessarily true, since in that case it appears we cannot consider what would have been the case had such a hypothesis been false.[23]

Nevertheless, appeal to the tracking condition seems to me a promising approach to confirmation and to the raven paradox because, unlike other well-known approaches, it does aver to a link between the observed and the unobserved. And here it seems to me that Hume was right: insofar as we take the observed as a guide to the unobserved, this is because we suppose that there is some connection that binds them together. From a tracking perspective, the solution to both the New Riddle of induction and the Raven Paradox depends on seeing that only some instances track others.

[22] Cf. D. Lewis, 'Counterfactual Dependence and Time's Arrow', in *Philosophical Papers*, Vol. II (New York: Oxford University Press, 1986), 32–66 .

[23] Cf. S. Kripke, *Naming and Necessity* (Oxford: Blackwell, 1980), Lecture III.

https://doi.org/10.1017/S1358246100009759 Published online by Cambridge University Press

And the difference between an arbitrary conjunction and a confirmable conjunction is that only in the confirmable case do some conjuncts track others.

The tracking condition has its origin in an account of knowledge rather than in an account of justification or confirmation. So insofar as this condition does indeed have a role to play in confirmation, perhaps we should see the aim of empirical testing of scientific hypotheses not just as the discovery of the truth, but also as the acquisition of knowledge. On this view, scientists seek through their empirical work to hook themselves up with nature in a certain counterfactual way. At the start of this essay, I distinguished between the justificatory and descriptive tasks in epistemology. I think the appeal to tracking is part of both. It is part of the descriptive task, where the raven paradox naturally appears, because I have suggested that we can account for our differential attitude towards the relevance of black ravens and white paper in part in terms of our judgement of a differential contribution to establishing a tracking relationship between evidence and hypothesis. But an appeal to tracking also forms part of a justificatory or normative account, since it is part of an account of what knowing that a hypothesis is true requires. So an appeal to a tracking condition has a place in an account of what it takes to have scientific knowledge, and we can improve our description of scientific practice by seeing inquirers as attempting to meet that condition.[24]

University of Cambridge

[24] I am grateful to audiences at Oxford University, Cambridge University, University of East Anglia, Leeds University, University of Aarhus, and the Royal Institute of Philosophy, and for helpful reactions to my talks about ravens, and to Alex Broadbent and Paul Dicken for comments on a draft paper.

https://doi.org/10.1017/S1358246100009759 Published online by Cambridge University Press