Paradoxes—Reading group 5

Chapter 5: Believing rationally

In this chapter, Sainsbury considers two *paradoxes of confirmation*: 'Goodman's paradox', and the 'ravens dilemma'. He the considers a separate puzzle in epistemology: the 'paradox of the surprise examination'.

Goodman's paradox

Hume's old problem of induction regards the question: 'Why should we think that hitherto-unobserved instances of some regularity in nature will be in line with previously-observed instances of said regularity?' For example: why think that the sun will rise tomorrow, having risen every day in the past? This kind of reasoning seems to underpin all scientific practice—but, as we know from General Philosophy, it is very hard to justify!

So-called 'new' problems of induction—such as *Goodman's paradox* and the *ravens paradox* (for the latter see below)—ask the question: *given* that it is at least sometimes justified to reason inductively, how do we establish what the good inductive inferences are supposed to be?

To see how this proceeds in the case of Goodman's paradox,[1] first define the predicates 'grue' and 'bleen' as follows:

> An object is *grue* iff it is green if first observed before time t, and blue if first observed after time t.
>
> An object is *bleen* iff it is blue if first observed before time t, and green if first observed after time t.

(Note that grue and bleen objects do not change colour—this is a common error!) Inductive inferences formulated using the predicates 'grue' and 'bleen' would seem to lead to incorrect conclusions being drawn—for example, suppose that I am walking through a forest just before time t; all trees I observe are grue, hence, I formulate the inductive inference, 'all trees are

[1] Articulated in his *Fact, Fiction, and Forecast*, ch. 3.

grue'. Time t then passes, and, because of my using the predicate 'grue', I will be *wrong* about the colour of any further trees that I observe after that time.

This in mind, Goodman's new riddle can be put as follows:

> **Why is it appropriate to form inductive inferences in terms of some predicates (e.g. 'green', 'blue'), but inappropriate to form inductive inferences in terms of some other predicates (e.g. 'grue', 'bleen')?**

Composite predicates

The standard response to the new riddle is to attempt to identify something defective with predicates such as 'grue' and 'bleen'. One first attempt at this proceeds by noting that predicates such as 'grue' and 'bleen' are (it would seem) *composite*, built out 'fundamental' predicates, such as 'green' and 'blue'. But this response will not do, because one could also define the predicates 'green' and 'blue' in terms of 'grue' and 'bleen':

> An object is *green* iff it is grue if first observed before time t, and bleen if first observed after time t.
>
> An object is *blue* iff it is bleen if first observed before time t, and grue if first observed after time t.

Thus, the situation is symmetrical—and some other means of excluding from application in inductive arguments predicates such as 'grue' and 'bleen' must be identified.

Swinburne's response

- Swinburne's response to the new riddle proceeds by first distinguishing *qualitative* from *locational* predicates—the latter are predicates which *must* be formulated with reference to space and time; the former are predicates which need not be so formulated.
- Swinburne correctly identifies standard colour predicates such as 'green' and 'blue' as qualitative, and non-standard colour predicates such as 'grue' and 'bleen' as locational. (To illustrate: around time t, users of the 'grue' predicate would have to look at their watches in order to talk about colours.)

- Why is this relevant? One does not need knowledge of temporal or locational facts in order to apply a qualitative predicate, whereas one *does* need such knowledge in order to apply a locational predicate.
- In light of this, Swinburne argues that on pragmatic, practical grounds, we are justified in using qualitative predicates over locational predicates.

Even if Swinburne succeeds in identifying an asymmetry between predicates such as 'green' and 'blue' on the one hand, and predicates such as 'grue' and 'bleen' on the other, one would be justified in feeling that merely identifying such an asymmetry is insufficient to solve the new riddle—for, while such a distinction may provide a means of identifying pathological locational predicates, and thereby avoiding their use, it remains to explain *why* such predicates lead to the formulation of bad inductive inferences.

Question: How can we refine Swinburne's account to address these concerns?

The ravens paradox

Let us now move on from Goodman's paradox, to consider a different paradox of confirmation—the *paradox of the ravens*. Here is how one might put this problem formally:

P1: Instances of an object a having as property P and a property Q confirm (i.e., raise our degree of belief in) the proposition 'All things which are P are Q'. That is, $Pa \wedge Qa$ confirms $\forall x\,(Px \rightarrow Qx)$. (*Nicod's condition*.)

P2: What confirms one proposition conforms any logically equivalent condition. (*Equivalence condition*.)

C: White shoes confirm 'all ravens are black'.

This conclusion appears, at first sight, to be paradoxical—for white shoes would seem to be *irrelevant* to whether all ravens are black. But, as one would expect, there have been many proposed solutions to the paradox. Here, I'll consider two of the most famous.

Natural kinds

Quine, in his paper *Natural kinds*, suggests that Goodman's paradox and the ravens paradox can be solved in the same way—by invoking the notion of a 'natural kind'. Natural kind predicates are meant to 'correspond to' nature—to 'carve nature at its joints' (to use an expression from Lewis). The idea is that only the predicates associated with natural kinds are 'projectable'—i.e., can form the basis for good inductive inferences. Then, Quine can say that 'grue' and 'bleen' do not correspond to natural kinds, and so are not projectable; similarly, 'non-black' and 'non-raven' are not projectable. (In the context of the ravens paradox, therefore, Quine rejects Nicod's condition.)

Question: Do you think a response to the new problems of induction invoking notions of natural kinds can succeed?

Bayesian confirmation theory

A different response to the ravens paradox—the *Bayesian solution*—accepts the (supposedly) paradoxical conclusion, but mitigates its force by introducing relative degrees of confirmation, and makes this quantitative using Bayes' theorem, which reads:

$$p(h|e) = \frac{p(e|h)p(h)}{p(e)}$$

Here:

1. $p(h)$ represents one's *prior probability* in hypothesis h—that is, it quantifies the probability that one assigns to h being true, before getting the new evidence e.

2. $p(h|e)$ represents one's *posterior probability* in hypothesis h—that is, it quantifies the probability that one assigns to h being true, after getting the new evidence e.

3. $p(e)$ represents one's probability about how likely it is to see evidence e.

4. $p(e|h)$ represents one's probability about how likely it is to see evidence e, *on the assumption* that hypothesis h is indeed true.

Example: drug tests

Suppose that a test for using a particular drug is 99% 'sensitive' and 99% 'specific'. That is, the test will produce 99% true positive results for drug users and 99% true negative results for non-drug users. Suppose that 0.5% of people are users of the drug. What is the probability that a randomly selected individual with a positive test is a drug user?

$$\begin{aligned} P(\text{User} \mid +) &= \frac{P(+ \mid \text{User})P(\text{User})}{P(+)} \\ &= \frac{P(+ \mid \text{User})P(\text{User})}{P(+ \mid \text{User})P(\text{User}) + P(+ \mid \text{Non-user})P(\text{Non-user})} \\ &= \frac{0.99 \times 0.005}{0.99 \times 0.005 + 0.01 \times 0.995} \\ &\approx 33.2\% \end{aligned}$$

Even if an individual tests positive, it is more likely that they do not use the drug than that they do. This is because the number of non-users is large compared to the number of users. The number of false positives outweighs the number of true positives. For example, if 1000 individuals are tested, there are expected to be 995 non-users and 5 users. From the 995 non-users, $0.01 \times 995 \approx 10$ false positives are expected. From the 5 users, $0.99 \times 5 \approx 5$ true positives are expected. Out of 15 positive results, only 5 are genuine.

The Bayesian interpretation

In the 'confirmation theory' understanding of Bayes' theorem which is of interest to us here, probability measures an agent's 'degree of belief'. Bayes' theorem then links the degree of belief in a hypothesis h before and after accounting for evidence e:

- $p(h)$, the *prior*, is the initial degree of belief in hypothesis h.
- $p(h|e)$, the *posterior*, is the degree of belief having accounted for e.

So, for example, we can interpret the above application of Bayes' theorem to the drugs case as follows. If initially we have a degree of belief of 0.5% in the person in question being a drug user (so we think it's very unlikely), and (given our knowledge of how sensitive and specific the test is) we then observe a positive test result for that person being a drug user, Bayes' theorem tells us that our degree of belief in that person being a drug user should increase to 33.2%.[2]

When looking at the ravens paradox, the Bayesian will give the following response: on reasonable assumptions about the relative numbers of ravens versus non-ravens in the world, seeing a black raven will, using the machinery of Bayes' theorem, confirm the hypothesis 'All ravens are black' much more than seeing a non-black non-raven. However, the latter evidence will also confirm the hypothesis 'All ravens are black' to some degree. Thus, the force of the paradox is mitigated, and the paradoxical conclusion is explained away.

Question: Can seeing a non-black non-raven ever confirm the hypothesis 'All ravens are black' *more* than seeing a black raven? (Answer: *yes*, given certain assumptions about the population one is dealing with.)

The paradox of the surprise examination

Finally, Sainsbury discusses the 'paradox of the surprise examination', which arises in epistemology, and which regards iterated knowledge. Suppose that a teacher announces to her pupils that she intends to give them a surprise examination at some point in the following term. The pupils can argue, as follows, that she will not be able to do this:

> *If you want the exam to be a surprise, then you cannot give it on the last day of term; for if you do, then we will know, on the second-to-last day, that it will be on the last day, and the exam won't be a surprise. You also cannot give the exam on the second-to-last day of term. For if you do, then we will know, on the third-to-last day, that it will be on either the last day or the second-to-last day, and will know, by the reasoning just described, that it will not be on the last day; so again the exam won't be a surprise. Parallel reasoning shows that you cannot give the exam on the third-to-last day, or the fourth-to-last day, or on any of the other days of term. Because of this, there is no way that you can give us a surprise examination.* (Internet Encyclopedia of Philosophy, The KK Principle, §3)

It is natural to think there must be something wrong with the pupils' reasoning; but it is hard

[2] It's interesting to think about how Bayesian confirmation theory bears upon the problem of induction: does it show that it's reasonable to believe that the sun will rise tomorrow? Or does it, rather, merely *codify* and *quantify* our inductive reasoning?

to see where the reasoning goes wrong. Many suggest that the issue here is with the pupils iterating their knowledge (cf. the 'KK principle'). Let's think about what the problem here is supposed to be, by considering the scenario in more detail. Here's how the IEP puts the issue:

> *Let part 1 of the pupils' reasoning be the part that rules out the last day, let part 2 be the part that rules out the second-to-last day, and so on. Since part 2 of the pupils' reasoning rests on the assumption that part 1 works, it is natural to say that part 2 works only if they know that part 1 works. And since part 3 rests on the assumption that part 2 works, it is natural to say that part 3 works only if they know that part 2 works, and thus, only if they are in a position to know that they know that part 1 works. Similar reasoning seems to show that part 4 works only if they are in a position to know that they know that they know that part 1 works, and so on. So the pupils' reasoning seems to assume that they are in a position to repeatedly iterate their knowledge of the fact that part 1 works, and it is not at all clear that this assumption is correct.*

Question: Do you agree that the assumption that knowledge can be iterated is the problem with the paradox of the surprise examination?