LANGARA COLLEGE

Philosophy 2201 – Theory of Knowledge

Richard Johns, February 2024

The Argument from Induction

1. The logical gap between evidence and theory

In our discussion of rationalism vs. empiricism, we've seen that the history of physics contains many arguments that are largely *a priori* in character. To the extent that these arguments are viewed as reasonable, cogent, or having epistemic weight, they provide some grounds for accepting rationalism. However, even if they are seen as having some force, they might yet be viewed as isolated cases that are superfluous to the scientific enterprise generally. Rationalists therefore try to show to that science cannot operate successfully without *a priori* knowledge in some form.

The best argument that rationalists have to support this claim is the *argument from induction*, which BonJour (*In Defense of Pure Reason*, p. 3) summarises as follows:

> "... if the conclusions of the inferences genuinely go beyond the content of direct experience, then it is impossible that those inferences could be entirely justified by appeal to that same experience. In this way, *a priori* justification may be seen to be essential if extremely severe forms of scepticism are to be avoided"

The key premise of this argument is that the propositional content of a scientific theory will always "go beyond the content of direct experience". It is clear that scientific theories (which describe unobserved entities such as the supercontinent of Pangea, black holes, quarks and Unruh radiation) are not logical consequences of any set of observations, for there is always more than one possible explanation of any given data set. Hume went much further than this, however, and claimed that a proposition that describes unobserved matters of fact must be *entirely independent* of one that describes our experience. Hume wrote for example:

> My experience directly and certainly informs me that that fire consumed coal then; but it's silent about the behaviour of the same fire a few minutes later, and about other fires at any time.

Experience, Hume says, is *silent* concerning facts that have not been observed, i.e. it gives zero information about them. In the absence of some background knowledge that connects the two together, it would then be illogical even to raise our epistemic probability that future fires will consume coal, based on past experience.

Hume is generally considered to be correct on this point, which can be illustrated by the case of predicting the contents of a cardboard box during a house move. Suppose you reach blindly into a box, and pull out a mug. What else is inside the box? We would expect that there are probably more kitchen items in there, but it's possible there would be a toothbrush, a can of paint, or even a live tortoise. Such inferences can be made using our background knowledge of how people usually pack boxes, but from an initial state of pure ignorance, observing one item would give you no information at all about the others.

This premise, that scientific theories are *trans-empirical*, is almost universally accepted among philosophers, being accepted of course by rationalists like Leibniz and Kant as well as empiricists like Hume and Quine. Leibniz noted for example that:

> The senses, although they are necessary for all our actual knowledge, are not sufficient to give us the whole of it, since the senses never give anything but instances, that is to say particular or individual truths. Now all the instances which confirm a general truth, however numerous they may be, are not sufficient to establish the universal necessity of this same truth, for it does not follow that what happened before will happen in the same way again.

As Quine summarises, "… one's theory of nature transcends any available evidence".[1]

This first premise is therefore rather uncontroversial. Nevertheless this premise, combined with two others that are also widely held, entails not just the existence but also the *indispensability* of a priori knowledge to the scientific enterprise. This 'argument from induction' can be summarised as follows.

1. In a scientific inference, the conclusion contains information not provided to us by sense experience.
2. In any rational inference, the information in the conclusion cannot go beyond the premises. (Reasoning cannot create information.)
3. Scientific inferences are rational.

∴ Scientific inferences require *a priori* knowledge.

Rationalist philosophers therefore see the trans-empirical nature of science as putting it in conflict with empiricism. They argue that, from a purely logical perspective, experience by itself cannot give any information at all about matters of fact that lie beyond experience. So, if the empiricists are right that all human

[1] This claim is also commonly referred to as "the under-determination of scientific theories by evidence".

knowledge comes from experience, then there cannot be any knowledge of scientific theories, but only of data.

Rationalists like Leibniz see *a priori* knowledge (i.e. innate knowledge, or knowledge prior to experience) as the solution here. Since scientific knowledge does exist, there must be some human knowledge that doesn't come from experience. Scientific theories do not logically follow from empirical data alone, so additional premises are needed, and these must be *a priori*. Rationalists say that *a priori* knowledge bridges the logical gap between data and hypotheses, or between appearance and reality.

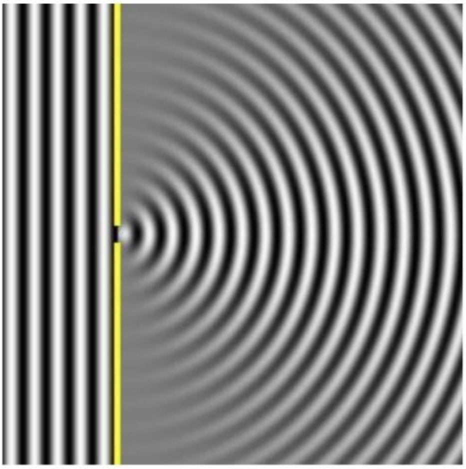
2. How do scientists *actually* infer theories from data?

If you ask philosophers of science today about the logical process by which theories are inferred from data, they will generally give one of two answers: inference to the best explanation, or Bayesian reasoning. The exact relationship between these two forms of inference is controversial, but here I will present them as complementary ways to analyse the same process, rather than as conflicting schemes. Both can be motivated by some early remarks by Christian Huygens, in his *Treatise on Light* (1678), where he describes a new "kind of demonstration" that is useful in science.

> One finds in this subject a kind of demonstration which does not carry with it so high a degree of certainty as that employed in geometry; and which differs distinctly from the method employed by geometers in that they prove their propositions by well-established and incontrovertible principles, while here *principles are tested by the inferences which are derivable from them*. The nature of the subject permits no other treatment. It is possible, however, in this way to establish a probability which is little short of certainty. This is the case when the consequences of the assumed principles are in

perfect accord with the observed phenomena, and especially when these verifications are very numerous; but above all when one employs the hypothesis to predict new phenomena and finds his expectations realized.

Huygens proposed that light consists of vibrations within an invisible elastic material (which he called the ‘ether’) that fills all of space. Thus, while of course our eyes detect light, the true nature of light (the movements of the particles of ether, the wavelength of the vibrations, the shape of the wavefronts, etc.) are invisible. This is quite different from ordinary mechanics, which concerns visible particles (billiard balls, planets, etc.) Consider for example the standard diagram below, showing plane waves approaching a solid yellow barrier with a small hole in it, so that the hole becomes a source of spherical waves. These wavefronts are of course not actually visible, so that (as Huygens says) we have to judge this model by what it predicts concerning things that we *can* observe.



Plane waves become spherical at the aperture

To see how this works, let H be some proposed scientific principle or law, such as Huygens' own Principle: *Every point on a wave front serves as a source of a spherical waves.* To test such a hypothesis, we logically deduce a prediction, or *observational consequence* from H, i.e. we show logically that if H is true then some observable event E should occur, when we do a certain experiment. (For example, Huygens' Principle entails that the behaviour shown in the image above will occur.) Then of course we do an experiment to see if E *actually* occurs. Huygens notes that if E is observed to occur, then this does not prove with certainty that H is true. "I do not think that we know anything very certainly but all probably." But if H predicts many separate events that agree with observation, and especially if some of those observations were previously unknown, then Huygens says that H can be very probable, indeed almost certain. Huygens' argument therefore has the following structure:

H predicts phenomena E_1, E_2 and E_3
E_1, E_2 and E_3 are observed to occur

$\therefore$ H is probably true

Students who have studied formal logic will notice that the above argument is an instance of "affirming the consequent", and is deductively invalid. This may not seem to be a problem, because the conclusion only says that H is *probably* true. But unfortunately there are cases where H predicts several observed data and yet is (intuitively) very unlikely to be true.

For example, let E_1, E_2 and E_3 specify the outcomes of three tosses of a coin, say *heads* in each case. Further, let H be the hypothesis that this particular coin *necessarily* lands heads every time (i.e. the chance of heads is one). We see that premise 1 is true here – in fact the prediction is certain, having probability one, since according to H the coin is *bound* to lands each time. Also suppose that premise 2 is true. So is the conclusion true as well? Is H (the claim that the coin always lands heads) probably true?

You might be reluctant to say that H is probably true. For one thing, there are *many alternatives* to H that also predict E_1, E_2 and E_3, although not with certainty. For example the coin might actually be fair, with an equal chance of heads and tails (0.5), and just happened to have landed heads on these three tosses. The chance of this occurring, 1/8, or 0.125, is not particularly small. Also, even if the coin *is* biased towards heads, it need not be as strong a bias as H claims. For example, if the chance of heads on each toss is 0.9 rather than 0.5, then the chance of 3 heads in 3 tosses is a very respectable 0.729. The hypothesis H is therefore just one possibility among many, so why should it possess most of the probability?

Another argument against H being probably true is the fact that it's hard to see how a coin could be made to land heads every time. The coin looks normal, let's suppose, with the Queen's head on one side and "tails" (no head) on the other. It seems rather unlikely that such a normal-looking coin could be sure to land heads on all tosses, doesn't it?

Notice how, in the previous paragraph, I said that H is "rather unlikely". This is an assignment of low probability, but what kind of probability is this? It is certainly a kind of epistemic probability rather than physical probability or chance, as H is a proposition that is already either true or false, not a future event that may or may not occur. Notice also that this (low) probability has nothing to do with the data E_1, E_2 and E_3 (three heads), for these data actually *support* H, and thus cannot reduce its probability. E_1, E_2

and E_3 are exactly what we should expect to get, if H were true. So this improbability of H has nothing to do with the data, and is therefore said to be "prior to" the data.

To remedy these defects with Huygens' scheme, we can add two more premises as follows:

1. If H is true then phenomena E_1, E_2 and E_3 are likely to occur,
2. E_1, E_2 and E_3 are observed to occur
3. H is likely to be true, *prior* to E_1, E_2 and E_3 occurring.
4. Alternatives to H are less probable, prior to the data, or assign lower probability to the phenomena.

$\therefore$ H is probably true

This is essentially Bayes' theorem, in a non-quantitative form. To make it quantitative, we have use the three kinds of "Bayesian probabilities": priors, likelihoods and posteriors.

Priors, e.g. P(H), measure the probability of each hypothesis *prior* to learning that E occurred.

Likelihoods, e.g. P(E | H), measure the degree to which E is predicted to occur by H. I.e. *assuming* H is true, it measures how likely E is to occur.

Posteriors, e.g. P(H | E), measure the probability of H being true *after* we learn that E occurred.

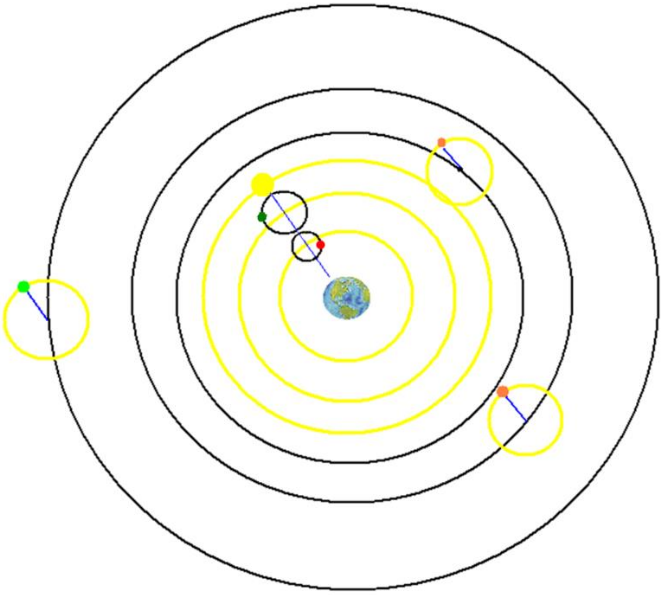
To calculate $P(H_1 \mid E)$, Bayes' theorem tells us to first multiply the likelihood of E for each hypothesis by its prior. Thus we calculate the products $P(H_1 \mid E)\times P(H_1)$, $P(H_2 \mid E)\times P(H_2)$, etc. We can call these products *Strength*(H_1), *Strength*(H_2), etc., noting that that the higher the value of $P(H_1 \mid E)\times P(H_1)$ is, the better or stronger H_1 is as an explanation of E, and the more posterior probability it deserves after learning E. Bayes' theorem can then be stated as follows, in the case of two possible hypotheses:

$$P(H_1|E) \quad = \quad \frac{Strength(H_1)}{Strength(H_1) \ + \ Strength(H_2)}.$$

In other words, Bayes' theorem assigns a posterior probability to each hypothesis in proportion to its strength, relative to the total strength of all the hypotheses added up. For example, suppose the priors of H_1 and H_2 are equal, but H_1 predicts E much more strongly than H_2 does. Then $P(H_1) = P(H_2) = 0.5$, but let $P(E \mid H_1) = 0.9$ and $P(E \mid H_2) = 0.05$ for example. In that case $Strength(H_1) = 0.45$, and $Strength(H_2) = 0.025$, so that $P(H_1 \mid E) = 0.45/(0.45 + 0.025) = 0.947$.

3. Copernicus argues for heliocentrism

To see how Bayesian reasoning applies to real science, let us examine Copernicus's main argument for a sun-centred universe. The argument is based on a very curious feature of Ptolemy's earth-centred model. (Note that this model had been the scientific orthodoxy for about 1,300 years). Ptolemy's model, shown below, has all seven planets (the moon, Mercury, Venus, the sun, Mars, Jupiter and Saturn) orbiting the earth. (The moon isn't shown in the diagram.) Apart from the sun and moon, all the planets move along 'epicycles', i.e. they orbit in a small circle around a point that itself orbits the earth. That's perhaps curious enough, but the *really* odd feature is that six of these circular motions are perfectly synchronised with each other. The synchronised orbits are the six yellow circles shown in the diagram below.



The sun (shown as a yellow disc) orbits right around the earth once per year, of course – that's what defines a year. Now, since Mercury and Venus are observed always to stay close to the sun, Ptolemy's model fixes the centres of their epicycles so that they always lie on the earth-sun line, as shown. With the 'superior' planets, those above the sun, you'll see that their epicycles are shown in yellow as well, since these little orbits also take exactly one year, and the position of the planet on its epicycle exactly matches the sun's position on its orbit, as shown. Ptolemy was forced to constrain the model this way by the observed fact that the superior planets undergo retrograde (backwards) motion when they're in maximal opposition to the sun. At that time they stop their usual easterly motion in the heavens, go backwards (i.e. west) for a short while, and then continue east again!

From a Ptolemaic perspective, there's no logical reason why so many orbits should be synchronised to the solar orbit. This feature of the model is therefore said to be *ad hoc*, which means that it is there for the sole purpose (*ad hoc* = 'for the purpose') of making the model fit the empirical data. A feature of a theory is *ad hoc*, in other words, to the extent that it is a free parameter within the theory, so that it can be adjusted to fit the empirical data.

The Copernican model, as we all know, placed the sun at the centre of the universe, and made the earth a planet – the third rock from the sun, orbiting between Venus and Mars. How did this explain the observed motions of the other planets? Well, the fact that Venus and Mercury always appear close to the sun is a direct consequence of the theory, since they're actually close to the sun! They dart around the sun, not the earth, and on very short leashes.

Now, what about the odd phenomenon of 'retrograde' motion? Why does that happen? And why does it *only* happen to a planet when it's on the opposite side of the heavens from the sun? According to Copernicus all the planets, including earth, orbit the sun in the same direction (eastward). But the earth is moving faster than the three higher planets, being closer to the sun than them, and will overtake each of them sometimes. During such overtaking events, the earth and (say) Mars are actually on the *same* side of the sun, so that from the earth's perspective Mars is *opposite* the sun. And as earth overtakes Mars, Mars appears to be going backwards, like a slow truck that you're passing on the highway.

In other words, the data that, for Ptolemy, require arbitrary, contingent features to be added to the model, are rationally necessary within the heliocentric model of Copernicus. Build any universe you like, consisting of planets orbiting a star. As long as you're on one of the planets in the middle, and the inner planets orbit at higher speeds than the outer ones, some of the other planets will always appear to be close to the sun, and the others will go retrograde while in opposition.

It should be stressed that, *empirically* speaking, Copernicus's model was no more accurate than Ptolemy's. The big advantage was that it was much less ad hoc. Of course even Copernicus's model was somewhat ad hoc, as all theories are. No theory can be determined by reasoning alone! For example, the speeds of the planets, and their orbital diameters, were ad hoc for Copernicus, among many other aspects. It's worth noting that later heliocentric models (e.g. due to Kepler, and then Newton) became progressively less ad hoc.

From a Bayesian perspective, the two hypotheses had equal likelihoods, since they predicted (approximately) the same data, at that time before telescopes were invented. The heliocentric model has a higher prior, however (according to Copernicus) due to its simplicity.

Let's pause for a moment here and ask how significant this advantage is, for the Copernican model. For some, this huge reduction in ad hoc-ness gives rise to a strong feeling that heliocentrism must be right. Thomas Kuhn, on the other hand, says this advantage is "largely an illusion", and merely "a propaganda victory". Kuhn says that Copernicus's model appealed primarily to a "limited and perhaps irrational subgroup of mathematical astronomers" with a certain aesthetic preference that he calls a "Neoplatonic ear for mathematical harmonies". How cogent is Copernicus's argument, in *your* opinion?

This argument for heliocentrism also provides a good illustration of how rationalistic arguments are not purely logical. Copernicus was a Catholic priest, and like all theists he was a creationist in the broad sense, i.e. he thought that the universe was engineered by a super-intellect, a master engineer. Suppose that such an engineer decided to build a universe according to the Ptolemaic blueprint. Would this be logically possible? Surely it would – while getting the orbits all synchronised perfectly might pose technical challenges, even human clockmakers have overcome similar difficulties, so a divine clockmaker could certainly pull it off. So why did Copernicus think that God didn't in fact do that? In *De Revolutionibus*, Book 1, Chapter 10, Copernicus wrote that in making his model "We thus follow Nature, who producing nothing in vain or superfluous often prefers to endow one cause with many effects." In other words he believed that the creator, a wise engineer, would use the neatest, simplest, most economical mechanism available. As with Leibniz, Copernicus's *a priori* judgements were based on the wisdom of the creator, who creates a rational and comprehensible world, rather than selecting one of the many logically possible worlds at random.

4. Scientific vs. instinctive induction

Leibniz, as noted above, believes that scientific conclusions are obtained using reasoning, from innate principles as well as observations. However, he also recognises a rudimentary kind of induction that even animals do. He describes animals as 'empirics', which means that they will notice simple patterns and instinctively expect them to continue, like Pavlov's dogs who salivate whenever they hear a bell. Unlike scientists, animals are not curious about the unobserved forces that gave rise to that past pattern; nor do they distinguish between patterns that occur of necessity and those that happen merely by chance, or due to temporary circumstances. Thus animal induction does not involve theoretical reasoning, and is rather unreliable. Leibniz writes:

> "This is how man's knowledge differs from that of beasts: beasts are sheer empirics and are guided entirely by instances. Men can come to know things by demonstrating them, whereas beasts, so far as we can tell, never manage to form necessary propositions. Their capacity to go from one thought to another is something lower than the reason that men have. The thought-to-thought sequences of beasts are just like those of simple empirics who maintain that what has happened once will happen again in a case that is similar in the respects that they have noticed, though that doesn't let them know whether the same reasons are at work."

For example, a chicken will expect to be fed when the farmer enters the shed in the morning, as has happened on a hundred previous mornings, without any curiosity as to why the farmer should go to such trouble. And an empiric will expect day to follow night, and night to follow day, without any thought of the earth's rotation or other theories about the unseen causes of this observed pattern. A scientist by contrast can *predict new phenomena*, prior to observation, such as the midnight sun during an Arctic summer, but this requires a rational understanding of the underlying causes.

It is remarkable that David Hume's examples of inductive inference, in the *Enquiry*, Section 4, are all cases of instinctive induction rather than induction by reasoning. He mentions, for example, expecting day to follow night, expecting the next flame to be hot, the next loaf of bread to be nourishing, and so on. He doesn't mention anything remotely similar to the arguments in my "Rationalism in Physics" notes, even though four of these five arguments (and many similar arguments by Newton, Huygens, Kepler, and of course Copernicus) predated Hume. By cherrypicking examples in this way, Hume apparently hopes that the reader will be gulled into accepting his extraordinary view, that scientific conclusions are produced by non-rational instincts.

5. David Hume: The instinct theory of science

As stated in Section 1 above, Hume agreed with Leibniz that scientific conclusions go beyond experience. How then does he avoid Leibniz's conclusion that science needs *a priori* knowledge in addition to experience? Hume takes the only option available to him, and denies Premise 3 in the argument above, i.e. he denies that our belief in scientific theories is justified by reasoning. As Hume puts it (in the Bennett translation) "the conclusions we draw from that experience are not based on reasoning or on any process of the understanding."

But if inductive inference isn't any sort of logical reasoning, then how does it work? As mentioned in the previous section, Hume says induction works by a kind of 'natural instinct', which he calls 'custom', or 'habit'. We do many things by our innate human nature, Hume says, such as to love those who help us, and hate those who deliberately harm us. In a similar way, he says, we naturally and instinctively believe that observed patterns will continue into the future.

> Custom, then, is the great guide of human life. It alone is what makes our experience useful to us, and makes us expect future sequences of events to be like ones that have appeared in the past. Without the influence of custom, we would be entirely ignorant of every matter

> of fact beyond what is immediately present to the memory and senses. (*Enquiry*, Section 5, Part 1)

In other words, scientists form conclusions by exactly the same process of induction that Leibniz attributes to animals. Bizarre as this view is, Hume provides arguments for it that we should examine. The main line of argument can be summarised as follows.

1. Knowledge of *unobserved* matters of fact is founded upon the relation of cause and effect.
2. Knowledge of cause and effect comes only from experience.
3. Experience tells us only about the appearances of things, not their causal powers
4. No one has ever shown us how to logically infer the causal powers of an object from its superficial appearance.

∴ We cannot rationally infer that the same sensible qualities are likely to be accompanied by the same causal powers

∴ Scientific conclusions are not based on reasoning

∴ Scientific conclusions are based on natural instinct

The key premise here is (2), for which Hume offers three arguments. Here are the arguments, in summary.

Argument A Causal laws of nature cannot be proved through reasoning, since they aren't logically necessary. For example, the negation of a law of nature is perfectly conceivable, and not at all contradictory.

Argument B Reasoning is obviously unable to predict the effects experiments with complex systems with hidden "springs and mechanisms". For example, we cannot predict the effect of feeding bread to a tiger, or dropping a spark onto gunpowder.

Argument C In simple experiments, such as two colliding billiard balls, Hume simply asserts (repeatedly, in a variety of ways) that reason is powerless to infer the effect of a given cause.

Argument A: We must grant Hume's premise that causal laws are not logically necessary. One can conceive of alternative laws, without any logical contradiction. But this argument is really attacking a straw person, since no prominent physicist (that I know of) has ever claimed that laws of nature are logically necessary. Copernicus for example regards a geocentric universe as logically possible, as we can consistently imagine such a messy and wasteful universe of epicycles, eccentrics and equants.

Argument B is not a very serious argument, as Hume himself recognises. We must agree with Hume that *a priori* reasoning alone does not inform us about the functioning of complex mechanisms, such as a tiger's digestive system, or provide us with detailed knowledge of chemical reactions. The purported innate principles used by physicists are obviously much more general and abstract than that.

Argument C occupies quite a lot of space in Part 1 of Section 4, but it doesn't have much substance. Here's a sample of what Hume says:

> If we are asked to say what the effects will be of some object, without consulting past experience of it, how can the mind go about doing this? It must invent or imagine some event as being the object's effect; and clearly this invention must be entirely

> arbitrary. The mind can't possibly find the effect in the supposed cause, however carefully we examine it, for the effect is totally different from the cause and therefore can never be discovered in it. Motion in the second billiard ball is a distinct event from motion in the first, and nothing in the first ball's motion even hints at motion in the second.

In this passage, Hume seems to simply assert his conclusion (that reason is powerless to discover the effect of a given cause) rather than giving us any evidence for it.

One odd thing about this supposed argument is that the 'collision problem' that Hume refers to is one that was carefully studied by physicists (such as Descartes, Huygens, and his student Leibniz) in the previous century. Moreover, the problem was initially solved (by Huygens) using deduction from rational principles[2] rather than by experiment – although the results agreed with experiments. Why does Hume say that reason has nothing to contribute to the collision problem, when reason has already solved it? Hume may have been unaware of Huygens' work, as he never mentions it, but he does give a general response to all reasoning of this sort (italics added):

> "Custom has such a great influence! At its strongest it not only hides our natural ignorance but even conceals itself: just because custom is so strongly at work, *we aren't aware of its being at work at all.*"

[2] See Huygens, "On the Motion of Bodies Resulting from Impact", completed in the 1650s but first published posthumously in 1703. Three of these general principles are: (i) Natural motion is rectilinear, at constant speed, (ii) Symmetry is conserved during collisions, and (iii) The laws of collision are the same in all uniformly-moving reference frames.

In other words, Hume claims that any reasoning like that of Huygens must implicitly be based on past experience of colliding balls, even though Huygens is unaware of the fact. It is certainly possible that the general principles Huygens' uses could be suggested by some experiments (a ball rolling on a pool table will move in a straight line, for example). But these principles can also be inferred from the rationality of the world, so that numerous and detailed experiments are apparently unnecessary.

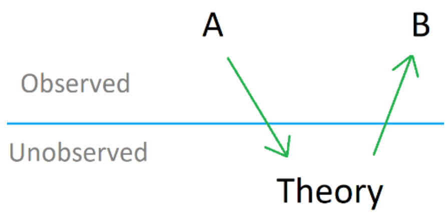
Hume's claim that scientific conclusions are always based on habit is also unable to account for *novel predictions*, which are common in the history of science. Consider again Newton's prediction that the earth is flatter in the Arctic than at the equator, which was actually contrary to the best observations at the time. Is it possible that Newton's calculations were irrelevant, since his belief *actually* arose from repeated past experiences of planets that were oblate spheroids, causing him to habitually associate the two ideas?

6. The thirst for (and confidence in) rational explanation

As discussed in the Sections 4 and 5 above, Hume errs in presenting science as the practice of inferring one observation statement from another. Science is rather the search for *explanations* of natural phenomena, i.e. the search for the (unobserved) *causes* of what we observe. Copernicus exemplifies the key assumption of science that patterns and coincidences *call for an explanation*. Rather than accept patterns (such as the six-fold repetition of the solar orbit) as ultimate facts, we should attempt to predict them from underlying causes.

The method called "inference to the best explanation" (IBE) is based on the deeply ingrained belief that there are deep and satisfying rational explanations for observed patterns. IBE is therefore biased, right from the start, against *ad hoc* explanations, or claims that a stable pattern is an ultimate fact, or that it is due merely to chance.

Scientists do frequently infer one observation statement from another, for example when they predict a future eclipse of the sun from past observations of the heavens. Such an inference however proceeds indirectly, via a theory, as shown in the diagram below.



Even though statements A and B are both observable, to infer one from the other requires a detour into the unobserved realm of theory. For example, if several balls drawn randomly from an unobserved urn are all observed to be black, then scientists don't *directly* come to believe that the next ball will be black. Instead, they first *explain* the observed black balls by the hypothesis that the urn contains all or mostly black balls, and then infer *from the hypothesis* that the next ball is likely to be black.

As I said in Section 2, the logical basis of inferring theories from data is provided by Bayes' theorem, but how does this relate to the method of IBE? To connect the two, we simply have to interpret the word 'best', in IBE, to mean *strongest,* in the sense of having the highest value of P(H | E)×P(H) (likelihood times prior). According to Bayes's theorem, the best explanation of E in this sense gets a larger share of the probability unit[3] than its competitors, and so in a probabilistic sense it is inferred from E.

The use of Bayes' theorem to make IBE precise and quantitative has the advantage of highlighting its dependence on assigning prior probabilities. In other words, IBE requires assessing the plausibility of each hypothesis, prior to learning the new observation E. Here an interesting question arises: Does Bayesian reasoning require (what we might call) *absolute priors,* i.e. assignments of epistemic probability to hypotheses in the total absence of empirical data? Or can the priors that science needs be rationally based on earlier observations? If the latter is true, then empiricism itself might survive the demise of Hume's instinct theory of science.

7. Bayesian empiricism?

Bayesian empiricism is the view that science is the attempt to describe the unobserved causes of our observations, and uses rational (Bayesian) inferences to justify those theories from the data. As an empiricist view, it is committed to the total absence of innate knowledge, so that the prior probabilities are always determined by prior experience alone. Unfortunately, such empirically-determined priors seem impossible in principle, as the following argument shows.

The argument involves what we might call *Goodman laws*,[4] which are say things like, "Newton's laws are followed up to March 7,

---

[3] Since competing hypothesis are mutually inconsistent, their probabilities sum to 1.
[4] Named after Nelson Goodman, who introduced the idea of such non-uniform laws in order to discuss inductive inference. See *Fact, Fiction and Forecast*, 1955.

2024, but after that <some other law> holds". The difficulty posed by such laws is that there is nothing wrong with them, from an empirical perspective, until the fateful moment of switchover. All their predictions are true up until that time, so it seems that they can be dismissed as improbable only on *a priori* grounds (e.g., to paraphrase Einstein, "God would be making a big mistake" in creating such a law). Yet if we assign the same prior probabilities to uniform laws as to their Goodman alternatives, inductive inference is impossible.

Is this argument too quick, however? Perhaps we can argue against Goodman laws on the empirical grounds that we have never found such a law to hold, in all our experience? Early scientists were perhaps merely lucky that their preference for uniform laws (i.e. non-Goodman laws) turned out well. But hundreds of years later, after seeing that uniform laws have a great track record of success, our preference for such laws has a strong empirical grounding, which is reflected in the prior probabilities we assign.

Unfortunately this response will not work. Hume himself considered a very similar argument, that the uniformity of nature should be inferred from the fact that it has followed such laws in the past. Hume pointed out that the argument is circular, since to project past uniformity into the future *assumes* the very uniformity in question. Our future is a realm that we haven't observed, and so we have no knowledge of it at all, if empiricism is true. To form beliefs about the future, based on observations about the past, would require knowledge that the past and future are at least likely to be similar in certain respects. But how could experience provide such knowledge, since our experience is entirely of the past?

8. Conclusion

The argument from induction is, as BonJour (1997, p. 3) notes, "extremely straightforward and obvious", but it is also barely mentioned by philosophers. Instead, philosophers talk of the *problem* of induction, and regard it a perennial difficulty that may never be solved. According to Leah Henderson in the *Stanford Encyclopedia of Philosophy*, for example, "a significant number [of philosophers] have embraced [Hume's] conclusion that it is insoluble."[5]

Other philosophers have not yet thrown in the towel, and Henderson's SEP article on the problem of induction presents a dozen different approaches to solving it, none of which is very similar to the position taken here. (The two closest approaches are those of Kant, and of Armstrong, BonJour and Foster.) Leibniz, who solved the problem correctly before Hume even posed it, is nowhere even mentioned. So I will let Leibniz have the final word:

> "… it is obvious that if some events can be foreseen before any test has been made of them, we must be contributing something from our side" (*New Essays on Human Understanding*, Preface)

[5] It's also noteworthy that Henderson's summary of Hume's argument does not even mention empiricism as a premise, even though this is absolutely essential to his conclusion. This is not at all unusual, and I see it as indicative of philosophers in general approaching this problem in an unfruitful way.