I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Graciela De Pierris, Co-Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

John Perry, Co-Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Mark Crimmins

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Thomas Ryckman

Approved for the Stanford University Committee on Graduate Studies.

Patricia J. Gumport, Vice Provost Graduate Education

This signature page was generated electronically upon submission of this dissertation in electronic format. An original signed hard copy of the signature page is on file in University Archives.
Abstract

In the first chapter, I explain “Hume’s problem of induction.” Induction cannot be justified *a priori*, and any argument that the unobserved will resemble the observed, based on observation, would be circular. I distinguish this problem from other forms of skepticism about induction, such as Goodman’s riddle, Hempel’s paradox, and other arguments for inductive skepticism made by Hume. I argue that the theories of Popper, Strawson, and Reichenbach fail to resolve Hume’s problem, arguing that Popper circularly assumes that nature is uniform and Strawson and Reichenbach do not claim, let alone defend, the intuitively indispensable truth that the sun will probably rise tomorrow.

In the second chapter, I discuss the closely related concepts ‘probability,’ ‘coincidence,’ and ‘explanation.’ I argue that probability needs no reductive definition, being widely understood and possibly primitive. Not all probability judgments are based on induction, so use of the concept ‘probability’ does not presuppose acceptance of induction. ‘Coincidence’ is identified with a certain sense of ‘unlikely’, and to ‘explain’, in the sense relevant to the topic of this dissertation, means to decrease the amount of coincidence we must accept.

In the third chapter, I discuss the family of solutions to Hume’s problem of induction that I endorse, in terms of inference to the best explanation, statistical sampling, and Bayesian reasoning. I summarize the arguments of Bayes/Price, Laplace, Mackie, Blackburn, Williams, Stove, Foster, Armstrong, Bonjour, and others who, I argue, share a common solution to the problem of induction (though they do
not all recognize this commonality). Strengths and weaknesses of the various formulations are discussed.

In the fourth chapter, I defend my version of the solution, an argument that the future will probably resemble the past which relies on a form of reasoning I call “inference to the only alternative to colossal coincidence” or “inference to lesser coincidence”. I argue that the only alternative to colossal coincidence is that the following principle holds true time-impartially: Regularities that have long persisted until a certain time are likely to continue somewhat further (in the absence of additional cross-inductive information). Colossal coincidence is *a priori* unlikely, so it is likely that this principle holds true time-impartially. I respond to several objections, including the possibility of time-restricted laws as an alternative explanation of past regularity, the possibility that we have been dealt a biased sample by our past experience, and the possibility that time-impartial dependence relations have either prior probability zero or non-existent prior probability, giving no ground to the Bayesian justification of induction.

I conclude with a discussion of the meaning of ‘causation’. The referent of ‘causation’ is whatever turns out to fill the role of preventing the vast regularities of our experience from being colossally coincidental, and has the property of being asymmetric with respect to the two directions of time (earlier-to-later and later-to-earlier). It is an open, empirical question, whether all physical dependence is causal.
Acknowledgments

I want to thank my advisor, John Perry, for his patience and focused criticism through numerous drafts. He helped me grow as a philosopher. It was a great honor to work with you, John.

To Graciela De Pierris, thank you for providing me with many new insights and helping me clarify my writing.

Thank you, Mark Crimmins, for your many great objections!

Thomas Ryckman, thank you for helping me gather my thoughts each time we spoke. You helped me identify additional material that enriched my work.

Ken Taylor, Krista Lawlor, David Hills, Alexi Burgess, and the many other faculty members who at one time or another contributed to the success of this work, thank you.

Thanks also to the staff and students of the Philosophy Department.

Thanks to my wife, Voke, my parents, and my brother, Rob, for their endless support and advice.

Last but not least, thanks to roommate of many years, Arvind Krishna, and my friends Akshan and Upal. You made Stanford more enjoyable!
# Table of Contents

Introduction .....................................................................................................................1

1. The Problem of Induction, and Other, Related Problems........................................13
   1.1 Hume’s Five Arguments for Inductive Skepticism ...................................................13
      1.1.1 Hume’s First Argument for Inductive Skepticism .............................................14
      1.1.2 Hume’s Second Argument for Inductive Skepticism .......................................18
      1.1.3 Hume’s Third Argument for Inductive Skepticism ...........................................18
      1.1.4 Hume’s Fourth Argument for Inductive Skepticism ......................................24
      1.1.5 Hume’s Fifth Argument for Inductive Skepticism ..........................................27
   1.2 Other Forms of Skepticism about Induction.............................................................29
      1.2.1 Deductivism ........................................................................................................29
      1.2.2 Skepticism in Philodemus and Goodman .........................................................30
      1.2.3 Hempel’s Paradox of the Raven .....................................................................35
      1.2.4 Cartesian Skepticism .........................................................................................38
   1.3 The Paradox of Induction .......................................................................................39
      1.3.1 Pyrrhonian Skepticism ......................................................................................41
      1.3.2 Popper, Strawson, Reichenbach .......................................................................42

2. Probability, Coincidence, and Explanation ................................................................51
   2.1 Probability ................................................................................................................51
      2.1.1 Introduction .........................................................................................................51
      2.1.2 The Definition of “Probability” .........................................................................53
      2.1.3 Epistemic and Objective Probability .................................................................60
      2.1.4 Criteria of Probability .........................................................................................66
      2.1.5 Quantitative Probability, and the Principle of Indifference ..............................68
      2.1.6 Probability and Cross-Induction .....................................................................77
   2.2 Coincidence ..........................................................................................................78
   2.3 Explanation ............................................................................................................86

3. Critical History of the Argument .............................................................................93
   3.1 Introduction .............................................................................................................93
   3.2 Bayes and Price .....................................................................................................94
   3.3 Laplace ...................................................................................................................97
   3.4 D.C. Williams .......................................................................................................99
   3.5 De Finetti and Mackie: Alternatives to the Principle of Indifference .....................100
   3.6 Blackburn ............................................................................................................104
   3.7 Foster, Armstrong, and Bonjour .........................................................................106
   3.8 Chapter Conclusion ..............................................................................................118

4. Solution, Objection and Replies .............................................................................119
   4.1 Solution ................................................................................................................119
   4.2 Objection 1: Fair sampling or Selection Bias? .....................................................130
   4.3 Objection 2: Bias Toward Marked or Unmarked States: Equally Likely? ..........134
   4.4 Objection 3: Zero Prior Probability of Dependence? ..........................................136

Conclusion: Induction and Causation ........................................................................142

Bibliography ..............................................................................................................148
Introduction

Most philosophers today hold a view that is implausible: Hume’s problem of induction cannot be solved.¹ If this view were correct, it would imply that we have no epistemic justification for believing that the Sun will rise tomorrow, and, indeed, imply that it is not the case that the Sun is likely to rise tomorrow. The predictions of astrologers and astronomers would be equally irrational, the results of pure guesswork. Induction would forever remain “the scandal of philosophy.” (Broad, p. 153)

I argue that a common solution to the problem of induction is successful: We understand intuitively that coincidence is unlikely, and are naturally drawn to accept its alternative (whether or not we know specifically which alternative, or can even fully imagine an alternative). I argue that this intuition can be used to justify induction in terms of a more primitive and more plausibly a priori form of reasoning I call inference to lesser coincidence.

Solutions that fall within this family of solution to Hume’s problem take various shapes with different surface features, such as Bayes’ Theorem (e.g. Richard Price), statistical sampling (e.g. D.C. Williams), and inference to the best explanation (e.g. John Foster). Many philosophers have defended these solutions convincingly, presenting arguments (all roughly the same, in spirit) that the future is likely to resemble the past. Still, the arguments can benefit from elucidation.

Appeals to inference to lesser coincidence to justify induction first appeared shortly after Hume first stated the problem (and has roots in much earlier arguments).

¹ Throughout this dissertation, I will use “the problem of induction” and “Hume’s problem” interchangeably, to mean Hume’s central argument for skepticism about induction in T1.3.6 and E4.2, i.e. that any argument that unobserved instances resemble observed instances on the basis of observation is circular. This will be developed in Chapter I.
From Thomas Bayes and Richard Price to Pierre Simon Laplace, and more recently D.C. Williams, David Stove, Roy Harrod, Simon Blackburn, J.L. Mackie, John Foster, David Armstrong, Laurence Bonjour, Scott Campbell and James Franklin, and Timothy McGrew, these philosophers reason (or can be interpreted as reasoning) that a coincidence is unlikely, raising the probability of the only alternative (or disjunction of all alternatives) to coincidence.

As I shall argue, any alternative to coincidence with respect to the observed regularities must render continuation of these regularities more likely. This “inference to the disjunction of alternatives to coincidence” form of reasoning is ubiquitous in philosophy, science, mathematics, and everyday life. (See Cartwright p. 75, and Owens p. 84) I will argue that it is pre-inductive. (See Ch. 2) The greater the observed correlation, the greater a coincidence it would be if there were no correct coincidence-lessening description (a particular kind of explanation), and so the more likely it is that there is a coincidence-lessening description to be discovered.

If the future were not likely to resemble the past, then past futures would not have been likely to resemble past pasts, I argue.² Therefore, given the repeated resemblance of “future” to past in the past, this past repetition would be colossally coincidental (as the product of so many low probabilities). This is extremely unlikely, so the alternative is rendered more likely. The present probability that the future resembles the past is the same as past probabilities that past futures would resemble past pasts. I shall return to defend this claim shortly. Colossal coincidence is too improbable to be believed, so the future is likely to resemble the past at any time up until which past futures have repeatedly resembled past pasts. In other words, the

² The language of “past pasts” and “past futures” is from Russell (1997).
future is likely to resemble the past because it would be an incredible coincidence if there were no time-transcendent/time-impartial likeliness for vast, stable regularities in progress to continue, given our observations of such vast, stable regularities in the past. Vast, stable regularities in progress must be time-impartially likely to continue, unlikely to cease at any particular time\(^3\), or else their existence in the past would be a colossal coincidence.

But perhaps this is not obvious: Why can only time-impartial probabilities serve as an alternative to colossal coincidence? Why not time-partial probabilities? To answer this question, consider an alternative to the time-transcendent probability hypothesis: a spatiotemporally limited, time-restricted law that grounds probabilities (See Foster, 70). Suppose, for example, that the law of gravitation is time-restricted. Given this, it may end exactly now. But even if gravity is time-restricted, let us ask, is it likely to stop at any particular time if it has long persisted until then? If the answer is “yes,” then it is a tremendous coincidence that it has not stopped for so long. Is it a coincidence that attosecond after attosecond, aeon after aeon, this time-restricted law has held steady?\(^4\) If so, it is a coincidence of colossal proportions. If a law is subject to change with time, ungrounded in a changeless meta-law, yet it persists for a long time, it is as colossally coincidental as any regularities that it might be invoked to explain. Vast, stable, time-restricted laws, in the absence of higher-order time-impartial laws “holding them in place,” are just as coincidental as first-order regularities with no coincidence-lessening explanation. Only time-impartial probability relations can

\(^3\) Phrasing suggested by Thomas Ryckman.

\(^4\) The history of the Universe can be carved into roughly \(4 \times 10^{17}\) seconds, or \(4 \times 10^{15}\) attoseconds. The question is why the Universe has maintained resemblance over so many instances when it could have changed.
provide the possibility of an alternative to colossal coincidence, given our experiences of vast and stable regularities.

Like Mackie and Bonjour, I endorse both the sampling and inference to the best explanation forms of the solution. Like Blackburn, I favor an epistemic reading of the argument, and focus on an epistemic notion of probability. (See Blackburn, 149-150. This will be explained in more detail in Chapters 2 and 3.) The following statement of the solution is meant to be intuitive, although not fully elaborated:

Premise (i) The vast regularities that we have observed, such as the Sun rising repeatedly, are colossally coincidental unless it is a time-impartial principle that vast and stable regularities in progress are epistemically very likely to continue to the next case; Premise (ii) Colossal coincidences are a priori unlikely; Therefore, Conclusion 1: Vast regularities in progress are, probably, time-impartially very likely to continue to the next case. Conclusion 2: It is more likely, given this conclusion, that the vast and stable regularities in our experience will continue to the next case.

For example, it is more likely that the Sun will continue to rise for some period of time, given that it has been observed to rise with such regularity in the past, than it would be in the absence of this past observed regularity.

This dissertation aims to clearly articulate both the problem and the solution, borrowing features of several accounts of the solution.

To set the stage for the solution, I believe we can help to clarify the problem of induction by isolating it from extraneous forms of skepticism about induction. This can be partly achieved by adopting an inductive premise that is idealized in several

---

5 More precisely, we must bracket both potential cross-inductive arguments (which can be assimilated into the overall inductive inference) and anthropic alternatives (according to which we are conscious of coincidence because coincidence is required for the existence of consciousness). To bracket cross-inductive arguments is to accept the requirement of total evidence of Hempel and Carnap (see Carnap, 213).
respects. We can strengthen the inductive premise without losing track of the problem. In the actual world, we observe various regularities and correlations: the Sun rises roughly 365 days each year, vegetables are beneficial to health, etc. On the other hand, we also experience deviations from regularity. From day to day we do different things and our environment changes. For the sake of argument, we can stipulate that these deviations do not exist, imagining that every day, our observed environment, thoughts, impressions, emotions and actions have been exactly the same for as long as we can remember.\(^6\) The problem of induction remains: What reason, if any, do we have to believe that our impressions are likely to continue going through the same daily routine tomorrow? For a long time, the actual world has been very nearly cyclical, but we can imagine a world that has been more fully cyclical, with every detail repeating exactly, and still raise Hume’s radically skeptical doubt as to whether the cycle is likely to continue any further.

Let us call this inductively idealized world, in which every day has always been exactly the same as the day before, “Groundhog Day.”\(^7\) For example, each day

\(^6\) Hume imagines such a world: “'Twou'd be very happy for men in the conduct of their lives and actions, were the same objects always conjoin'd together, and, we had nothing to fear but the mistakes of our own judgment, without having any reason to apprehend the uncertainty of nature.” (Hume, T1.3.12) In such a world, all of our experience would fall within what Hume calls “inductive proof.” (Hume, T1.3.11) Similarly, Carnap asks us to imagine, for the sake of inductive logic, an observer with a “simplified biography... whose entire wealth of experience is so limited that it can easily be formulated and taken as a basis for inductive procedures.” (Carnap, 213) He argues that such deviations from reality do not make inductive logic irrelevant for science or everyday life, any more than the theoretical idealizations of geometry, physics and chemistry (e.g. circles, frictionless motion, and ideal gases) make these studies irrelevant. (Carnap, 213-214)

\(^7\) Groundhog Day is named after the movie starring Bill Murray (Columbia Pictures, 1993). In the movie, unlike this thought experiment, Murray’s character’s thoughts and behavior change from day to day, influencing the otherwise repetitive world around him. The inductive premise is stronger in the movie than reality, but stronger still in this thought experiment.

We can consider two versions of Groundhog Day, one in which the identity of observations from day to day is known to go back infinitely into the past, another in which it is merely known to go back for a very long time, at least as long as anyone knows. Imagine that the world has a beginning. On the first day, we have certain observations. On the second day, we have the same observations, except that we now also observe our memory of having the same observations the previous day. On the third
at 3:15 pm GMT in Groundhog Day, the same 184 people are thinking about Hume’s problem of induction, just as they were the day before at the same time. The problem of induction arises with the same force in Groundhog Day as in the actual world, so a solution within its controlled parameters would solve the problem in principle.

In the same spirit of focusing the problem, let us also set out to defend only the weakest “principle of uniformity” needed to solve the problem of induction. First, the principle can be conditional. Second, we do not need to justify certain or universal conclusions; probable, singular inductive inferences will suffice. Our principle need not be any stronger than that vast regularities in progress are time-impartially likely to continue, *ceteris paribus.*

---

8 “Principle of uniformity” has conditional and non-conditional senses.

Conditional: For example, J.L. Mackie states the conditional uniformity principle “that individual things and goings-on tend to persist as they have been.” (Mackie, 1979, 170) This presupposes the condition that things have regularly been a certain way. If no regularity/pattern has occurred, for the world to persist as it has been is, seemingly “paradoxically”, to persist in a lack of regularity. P.F. Strawson explains this as a “higher-order induction”: “where irregularity is the rule, expect further irregularities.” (Strawson 1963, 262) The principle to be discussed, that vast regularities are time-impartially likely to continue, is a principle of uniformity in only the conditional sense.

Non-conditional: In contrast, Hume’s principle of uniformity is non-conditional: “Instances, of which we have had no experience, must resemble those, of which we have had experience, and... the course of nature continues always uniformly the same.” (Hume, T1.3.6) Similarly, Roy Harrod uses the expression “uniformity axiom” in the non-conditional sense when he argues that “if we are to make any progress out of our initial state of nescience... there must be some mode of induction that is independent of any uniformity axiom.” (Harrod, 11) As an example of a uniformity axiom that needs no such condition, Harrod writes of “a world of events or bodies which showed no uniformity whatever.” (Harrod, 10) To fill this out, we might imagine that there are only momentary appearances with no statistical bias toward sameness over time. Harrod’s reasoning is that since a lack of uniformity is possible, uniformity cannot be known to hold *a priori.* Here Harrod assumes that no uniformity in the sense of no stable relative frequencies means no principle of uniformity in any sense. But this disregards the possibility of allowing for a conditional principle of uniformity, whose condition is not satisfied in a world in which no stable relative frequencies are manifested (and thus this world cannot serve as a counter-instance). Harrod might be implicitly referring to this distinction when he calls the principle of uniformity “a little woolly.” (Harrod, 10)

9 By “ceteris paribus,” I mean, “in the absence of additional information justifying an overriding cross-induction.” In other words, all of our evidence must be taken into account. In Groundhog Day, we can,
Therefore, instead of setting out to justify a non-conditional, deterministic, universal, Newtonian principle of uniformity, we should set out to justify only a conditional, probabilistic, singular principle of uniformity. In particular: For any world at any time, given only (the condition, i.e. the inductive premise) that vast, persistent regularities have been observed until that time, it is unlikely that the regularities end at exactly that time. Justifying this principle would refute the Humean view that no argument can establish the probabilistic relevance of observed data to unobserved data, and would therefore solve the core of the problem.

As I shall explain, the principle that vast regularities are time-impartially likely to continue can be derived from the simpler, analytic principle that *epistemic probability supervenes on psychological structure* (i.e. intrinsic features of experience, narrow mental content, or Fregean *Sinn*). Intuitively, if two individuals are psychologically identical, their beliefs are subject to the same epistemic probability norms, and they encounter the same relative epistemic probabilities. For example, my doppelgänger on twin Earth and I share the same epistemic probability that our respective Sun will rise tomorrow. Like exchangeability in Bruno de Finetti’s “intrinsic characterization of induction,” the principle that epistemic probability supervenes on psychological structure equalizes probabilities over time, enabling us to derive present probabilities from past probabilities. (De Finetti, 2008, p. 78. I explain this claim in chapter 3.)

The epistemic probabilities that we should assign to various possibilities depend on (specifically, supervene on) how things *seem* to us. If everything seems

theoretically, include all of our past and present experiences in our inductive premise when asking whether, for example, the Sun is likely to rise tomorrow, so the ceteris paribus clause is satisfied.
exactly the same to two individuals (or one individual at different times), the same epistemic rationality constraints apply to each of them. For instance, if things seem the same to Jim today as they did yesterday, it cannot have been irrational for him to believe yesterday that the regularities would probably end at that time, yet rational for him to believe today that the regularities will probably end now. The identity of his epistemic landscape at the two moments creates a symmetry, which precludes the possibility of such distinctions in epistemic probability. Because all of his ideas and impressions are the same, his epistemic probabilities are also the same.

Combining the supervenience of epistemic probability on psychological structure with the ideal inductive premise of Groundhog Day, we can justify the conditional, probabilistic, singular uniformity principle that vast, persistent regularities in progress are likely to continue somewhat further. Given the identity of past and present psychological states, the same epistemic probabilities hold. If vast, persistent regularities were not likely to continue in the past, Jim’s observations of their repeated continuations in the past would be colossally coincidental, which is implausible. Thus they must have been likely to continue in the past. Given that Jim’s psychology now is exactly the same as it was in the past, and given that probability supervenes on psychological structure, vast regularities must be equally likely to continue from Jim’s present point of view. Supervenience of probability on psychological structure can serve as the ground of non-circular solution because it allows us to abstract from particular cases, transcending the observed/unobserved divide. (That is, if A

---

10 However, this does not mean that the two psychologically indiscernible Jims can’t differ in that what one is led to infer is true, while what the other is led to infer is false.

11 This must be set relative to his environment, including temporal location. At each day at the same time, it is the probability that the Sun will rise “tomorrow” that is in question, but the referent of “tomorrow”, of course, changes.
supervenes on B, this supervenience relation holds in all instances at all times, by
definition.)

It is a coincidence if an evenly weighted coin lands heads ten times in a row by
chance. It is unlikely to happen often, but likely to happen occasionally. However, it is
a much greater, colossal, unbelievable coincidence if an evenly weighted coin lands
heads 100 times in a row by chance. It will almost certainly never happen. Hume
recognizes this principle in his discussion of the constant conjunction between
impressions and their corresponding ideas (their copies): “Such a constant
conjunction, in such an infinite number of instances, can never arise from chance; but
clearly proves a dependence of the impressions on the ideas, or of the ideas on the
impressions.” (Hume T1.1.1) Given repeated observations of something, the more
coincidental these observations would be if due to chance, the less likely it is that they
are due to chance.

We believe the existential proposition that there is some explanans to be
discovered, such as an uneven weighting of the coin or a magnetic force, even if we
have not yet figured out what this explanans is. Similarly, facing the problem of
induction, we believe that there is some explanans, known or unknown, of our
observations of repetitiveness in the past, and use this inference to argue that the
repetitiveness is likely to continue somewhat further. In the case of the problem of
induction, however, the following additional argument is needed in order to infer that
the repetitiveness is likely to continue, based on the inference to the explanans:
Whatever the explanans might be, it must be less coincidental than the explanandum
would be in its absence, if it is to provide us with any reason, based on our
observations, to believe that it holds/obtains/exists.

If we explain the repetitiveness of our experience of gravity by reference to a time-restricted law, then either this time-restricted law itself is likely to continue once in progress or it is a colossal coincidence that it has persisted throughout all of our past and present experience. If this time-restricted law is likely to continue because of a higher-order time-restricted law, then either this higher-order time-restricted law is likely to continue once in progress or its past continuation is a colossal coincidence. The appeal to iterations of higher-order time-restricted laws never begins to reduce coincidence unless an unrestricted law with a probability to persist at any time is eventually reached. If an unrestricted law is not reached and the persistence of the time-restricted laws is a colossal coincidence, then these time-restricted “laws” do not play the relevant type of explanatory/coincidence-reducing role, so there is no longer any reason to believe that they exist.

The inference from past probabilities to present probabilities via this supervenience principle is most straightforward in the Groundhog Day world, where our experiences have always been cyclical, i.e. repeating each day. In Groundhog Day, where our psychological state has been the same each morning at the same time, the principle that epistemic probability supervenes on psychological structure entails that our epistemic probabilities, relative to our respective futures, have also been the same each morning. For example, for everyone living in Groundhog Day at any time, the probability, today, that the Sun will rise tomorrow, is the same as the probability, yesterday at the same time, that the Sun would rise today (aka yesterday’s tomorrow), and so on for two days ago, three days ago, etc. Unless it is highly likely, each day,
that the Sun will rise the next day, it is a colossal coincidence that it rose with such
great regularity in the past. A colossal coincidence is *a priori* extremely unlikely. This
raises the probability that the Sun will rise tomorrow.

Only a time-impartial likeliness for patterns in progress to persist could serve
as an alternative to colossal coincidence, given the vast regularities observed. For
laws, powers, causal connections, regularities, or anything else, if they are as likely to
cease as to continue at any time, having no time-impartial probability to persist once in
progress, they are unlikely to hold for long, and the fact that they have held in the past
can explain nothing because their own past continuation over time is just as
incredible.\(^\text{12}\)

Because the principle that epistemic probability supervenes on psychological
structure/appearance (uninfluenced by unobserved extrinsic features of my
environment) is not subject to change with time, it can provide the time-impartiality
needed to overcome the threat of circularity. This time-impartial likeliness might be a
higher-order “law” than the laws of physics, which might be local and subject to
change. In contrast, the fact that a regularity which has held persistently until now is
likely to hold in the next case, if the epistemic landscape now is the same as it was in
those past instances, is not subject to change. For example, the laws of gravity may
cease to hold, but probably not now, nor at any particular time after they have long
persisted. Only if something (such as a law or causal power) is time-impartially likely
to continue, unlikely to cease at any particular time, can it provide an alternative to

\(^{12}\) In contrast, it is no coincidence that 4+3=7 every time the sum is calculated, because mathematical
relations transcend time.
colossal coincidence, because only then is its own persistence any less coincidental
than the fact that it allegedly “explains”.
Chapter 1. The Problem of Induction, and Other, Related Problems

The problem of induction is the paradox resulting from the charge that any argument that unobserved instances probably resemble observed instances, based on observation, must circularly assume that unobserved instances probably resemble observed instances. It is easy to mistake Hume’s problem of induction for distinct skeptical arguments regarding induction. Stove writes that many philosophers “father on Hume the sprouts of their own brains, especially where these are of a ‘sceptical’ tendency.” (Stove, 193) Those attempting to solve Hume’s problem should not be expected to hit a moving target.

There are many philosophical “problems” of induction (as well as many non-philosophical problems of induction), with complicated relations among them. Many philosophers posed skeptical challenges to induction before Hume. In order to solve the problem of induction, it is not necessary to respond to all forms of skepticism about induction. While this may seem obvious, it is frequently forgotten. It is irrelevant to object to a solution to the problem of induction by invoking ancient forms of skepticism, Cartesian skepticism, Goodman’s new riddle of induction, or other forms of skepticism that are not the same as Hume’s central problem. The problem of induction should hold its ground when these other problems are bracketed, if it is to remain a paradox in its own right.

1.1 Hume’s Five Arguments for Inductive Skepticism

Hume himself raises at least five hurdles to the justification of induction. Only one of these (#1 below) can be considered the problem of induction, though some of the others (#2,3,4 below) play a supporting role, helping to set it up (#2,3), or reinforce it
by constraining possible answers (#3,4). The final hurdle (#5) seems to be a separate skeptical objection to induction, which applies equally to deduction. (See De Pierris 2005, p. 107)

1.1.1 Hume’s First Argument for Inductive Skepticism (i.e. “The Problem of Induction”)

Hume recognizes three potential sources of justification for induction: intuition, demonstration, and experience. Intuition involves only a single step of immediate recognition. For example, we see two stop signs and immediately recognize that they resemble each other with respect to color, and name this common color “red”.13 Demonstration involves multistep reasoning, such as proofs in mathematics. Intuition and demonstration, unlike experience, are classified by Hume as a priori, i.e. “relations of ideas.” (Hume E4.1) Hume asks which, if any of the three (intuition, demonstration, or experience), is the rational basis of inductive inference. Hume claims that we do not know by intuition or demonstration that the future will resemble the past, assuming that a priori knowledge of the correctness of inductive reasoning would falsely imply contradiction in the idea of the unobserved future not resembling the observed past. (T1.3.6, p. 89)14

---

13 Hume writes that resemblance, as well as contrariety and degrees in quality, “are discoverable at first sight, and fall more properly under the province of intuition than demonstration.” (Hume T1.3.1) De Pierris explains, “I can determine solely on the basis of an intrinsic content now ostensively present, prior to experiencing any other particular content, which intrinsic features a new experience would have to have in order, for example, to resemble the one now before the mind.” Resemblance relations between ideas do not change as long as the ideas themselves do not change, and so they are “certain, necessary, and a priori.” (De Pierris 2005, p. 104)

14 But it is contradiction in the idea of the unobserved future not probably resembling the observed past that we should be interested in. It is not clear that this idea is not contradictory; at least, it cannot be shown merely by pointing out that we can conceive the future not resembling the past, given that the probability that the future will resemble the past is logically consistent with the possibility that it will not.
With intuition and demonstration eliminated, Hume raises the circularity objection to any argument for induction based on experience. This circularity objection is the main part of the problem of induction. Any conclusion about the future based on past uniform experience must assume the principle that the future will resemble the past. This principle cannot be justified on the basis of an argument from past uniform experience, on pain of circularity. Hume writes, “It is impossible... that any arguments from experience can prove this resemblance of the past to the future: since all these arguments are founded on the supposition of that resemblance.” (E4.2, 37-38) Russell explains, “All arguments which, on the basis of experience, argue as to the future or the unexperienced parts of the past or present, assume the inductive principle.” (Russell 1997) This provides a template for objecting to a number of attempts to justify induction that are all shown to be circular in the same way.

Many proposed solutions to the problem of induction appeal to an explanans for the regularities observed, in order to establish the probability of unobserved regularities. Whatever explanans is offered, whether powers, forces, causal connections, laws, dependence relations, space-time worms, God, identity, or epistemic probability relations (and even if the account is ontologically noncommittal and the explanans is merely referred to as such), the skeptic can always find the Humean radically skeptical reply by demanding justification for the belief that the inferred explanans persists in the unexperienced parts of reality.

---

15 Graciela De Pierris helped with this formulation.
16 Here I use the term “explanans” instead of “explanation” to encourage the interpretation of it as an object (logical, physical, or otherwise) rather than as a token linguistic act of an explainer.
17 As I shall argue, the only way to block this objection is to argue that something timeless/tenseless, not subject to change, fills the role of explanans. The idea of a timeless explanans might seem queer, if not for the fact that many less controversial relations, including mathematical and deductive relations, are commonly regarded as not being subject to change with time or place.
To illustrate, suppose that we make an inference, based on past experience of watching billiard balls collide, that a transfer of energy or power is responsible for their behavior upon contact. Hume imagines an attempted solution based on this, which he rejects. “The past production implies a power,” and “The power implies a new production.” (Hume, T1.3.6, 91) Hume points out that this fallaciously assumes that the powers that we infer existed in the past will continue to exist in the future:

But it having been already prov’d, that the power lies not in the sensible qualities of the cause; and there being nothing but the sensible [observable] qualities present to us; I ask, why in other instances you presume that the same power still exists, merely upon the appearance of these qualities? (Hume, T1.3.6, 91)

We never directly observe this supposed power. We might object that we therefore have no reason to believe it ever existed. But Hume’s argument is different: he argues that even if we accept that our observations imply the existence of a power in the past, we may still have no reason to believe that the power continues to exist hereafter. All of our impressions of regularity are in the past and present, so it seems that anything we could infer from this regularity must also be confined to the past and present, including the power or probability. In sum, first, the skeptic asks what reason there is to believe that the objects of experience will continue to, for instance, communicate motion and follow the laws of gravitation. The anti-skeptic replies that the powers that have made objects interact as they have will continue to have the same influence. Second, the inductive skeptic asks why we should believe that these powers will continue to exist.

Russell describes a dialectic with a similar structure:

It has been argued that we have reason to know that the future will resemble the past, because what was the future has constantly become the past, and has always been
found to resemble the past, so that we really have experience of the future, namely of times which were formerly future, which we may call past futures. But such an argument really begs the very question at issue. We have experience of past futures, but not of future futures, and the question is: Will future futures resemble past futures? This question is not to be answered by an argument which starts from past futures alone. (Russell 1997)

Many people, the first time they are asked to defend the commonsense view that the future will probably resemble the past, reply that in the past, “the future” (what was then the future) has always turned out to resemble “the past” (what was then the past). Russell applies Hume’s circularity charge, asking what reason there is to believe that future futures will resemble past futures, just as Hume asks what reason there is to believe that future powers will resemble past powers (including believing that they exist).

We claim to know that the future will resemble the past because of our observations of the past. We think that the regularity of the past implies that there will be regularity between past and future. If there were no regularity in the past, we would not expect regularity to appear suddenly from nowhere; ironically, that would be a (higher order) form of irregularity (with respect to whether the world is regular). But if we think that the future will resemble the past because in the past, “the future” has regularly resembled the past, we are already assuming that the future will resemble the past in that higher-order respect. Thus, says Russell, our reasoning is circular.18

We do not know intuitively or demonstratively what the future will be like, writes Hume, and any “probable argument” that the future will be like the past based on experience will be “going in a circle,” because it must assume that the unobserved

18 Eight years later, Keynes writes rebelliously, “It is because there has been so much repetition and uniformity in our experience that we place great confidence in it. To this extent the popular opinion that Induction depends upon experience for its validity is justified and does not involve a circular argument.” (Keynes, 260)
resembles the observed with respect to whatever we appeal to in order to argue that they will be alike. (Hume, E4.2, p. 36) The challenge to show that the unobserved will resemble the observed in some particular respect becomes the challenge to show that the unobserved will resemble the observed with respect to whatever allegedly explains the regularity. For example, appealing to forces as a reason to believe that the Earth will continue to orbit the Sun begs the question by assuming that those forces will continue to exist, or operate in the way they have operated so far.

1.1.2. Hume’s Second Argument for Inductive Skepticism

Simply by soliciting a justification for induction, Hume creates a presumption that induction requires a defense in order to be justified. This is Hume’s “negative argument.” (Hume, E4.2.30, p. 34) He writes, “There is required a medium, which may enable the mind to draw such an inference, if indeed it be drawn by reasoning and argument... and it is incumbent on those to produce it, who assert, that it really exists.” (Hume, E4.2.29, p. 34) By challenging us to produce the argument for induction, Hume sets the stage for the circularity objection.

1.1.3. Hume’s Third Argument for Inductive Skepticism

Hume withholds the conceptual ingredients needed for a solution, challenging the meaning of our words with his restrictive theory of ideas. Hume’s theory of ideas does not allow us to use the ideas of epistemic probability and power/necessary connection that would be needed to solve the problem of induction. However, this theory is not

---

In a footnote to his main circularity argument in Enquiry 4.2, Hume writes, “The word, Power, is here used in a loose and popular sense. The more accurate explication of it would give additional evidence to this argument.” Hume sees his theory of ideas as providing “additional evidence” for his skeptical case. He deliberately separates the arguments, recognizing that his main circularity argument does not rely on his remarks “concerning the idea we form of power and efficacy”: “As such a method of proceeding may seem either to weaken my system, by resting one part of it on another, or to breed a confusion in
Hume writes: “Every idea is copied from some preceding impression or sentiment; and where we cannot find any single impression, we may be certain that there is no idea” (E7.2.61, p. 78); “All ideas are deriv’d from, and represent impressions” (T1.3.14); “Our ideas are images of our impressions” (T1.1.1); “All ideas are derived from impressions, and are nothing but copies and representations of them” (T1.1.7). Moreover, “we can never have reason to believe that any object exists, of which we cannot form an idea” (T1.3.14, p. 222). If we cannot trace our idea of an object to an impression or set of impressions, we can have no reason to believe that the object exists, says Hume.

Hume claims that Locke’s notion of power or necessary connection “is simply unintelligible; for it is a new idea that does not have a corresponding simple impression in the alleged case of single causation.” (De Pierris 2006, 298) De Pierris calls the idea of necessary connection, “The essential ingredient in our idea of causation.” (De Pierris 2002, p. 512) She writes, “The idea of necessary connection we in fact have... goes beyond any inspectable features before the mind and thus does not fulfill the justificatory demands of the theory of ideas.” (De Pierris 2002, 530) Hume writes: “We never have any impression, that contains any power or efficacy. We never therefore have any idea of power.” (1.3.14) Softening this only slightly, Hume concludes, “Either we have no idea of necessity, or necessity is nothing but that determination of the thought to pass from causes to effects.” (T1.3.14) “Necessity, then... is nothing but an internal impression of the mind.” (T1.3.14) It “exists in the my reasoning, I shall endeavour to maintain my present assertion without any such assistance.” (Hume, T1.3.6)
mind, not in objects.” (T1.3.14) This is less than what we ordinarily think we mean by “causal necessity,” namely something that exists beyond the mind, making us likely to succeed in the predictions we make based upon our belief in it.

Hume applies the same razor to our idea of probability:

What is here meant by likelihood and probability? The likelihood and probability of chances is a superior number of equal chances; and consequently when we say it is likely the event will fall on the side, which is superior, rather than on the inferior, we do no more than affirm, that where there is a superior number of chances there is actually a superior, and where there is an inferior there is an inferior; which are identical propositions, and of no consequence. (T1.3.11)

He finds the definition of probability in the criteria used to judge it. We judge that a symmetrical die is more likely to land on the disjunction of 1 through 5 than 6, because of the superior number of chances. Hume concludes that what we mean by probability in this case is nothing more than this condition or criterion. Probability just is the relative number of chances; this is not merely evidence of probability or one way of instantiating probability. Just as claims about powers, for Hume, are either meaningless or mere psychological descriptions, claims about probability are either meaningless, or mere descriptions of the criteria used to judge probability, thus divorced from rational credence.

Hume’s argument for the principle that ideas are derived from impressions is based on our experience of their constant conjunction, not an a priori argument. “If any one should deny this universal resemblance, I know no way of convincing him,

20 Puzzlingly, Hume says the same of mathematical necessity: “Thus as the necessity, which makes two times two equal to four, or three angles of a triangle equal to two right ones, lies only in the act of the understanding, by which - we consider and compare these ideas; in like manner the necessity or power, which unites causes and effects, lies in the determination of the mind to pass from the one to the other.” (T1.3.14)

21 But probability is not always found by comparing the number of chances, so Hume’s notion of probability seems fractured (though he tries to unify it). See Chapter 2 for further discussion of Hume’s theory of probability.
but by desiring him to shew a simple impression, that has not a correspondent idea, or a simple idea, that has not a correspondent impression.” (Hume T1.1.1) Hume even admits an exception to this rule: “There is however one contradictory phaenomenon, which may prove, that 'tis not absolutely impossible for ideas to go before their correspondent impressions.” (Hume, T1.1.1) In particular, Hume claims that we can have an idea of shades of blue that we have never seen between shades of blue that we have. “This may serve as a proof, that the simple ideas are not always derived from the correspondent impressions; tho' the instance is so particular and singular, that 'tis scarce worth our observing, and does not merit that for it alone we should alter our general maxim.” (Hume T1.1.1) But a single counterexample disproves the rule, and makes other potential counterexamples seem more likely.

For example, Thomas Reid argues that our ideas of number, space, duration, resemblance, identity, and existence “cannot be... called copys of any Impression either of Sensation or Reflexion,” and so “There seems ground to apprehend that we have many Ideas which are not copys of Impressions.” (Reid, p. 294) Hume, of course, disagrees. For example, he claims that we have impressions of space: “The idea of space is convey'd to the mind by two senses, the sight and touch.” (T1.2.3) The idea of space for Hume “is nothing but the idea of visible or tangible points distributed in a certain order.” (T1.2.5) But what is this “order,” and what simple impression does it correspond to? The matter is not beyond dispute: it is, at least, not as clear that we have impressions of space as it is that we have impressions of color.

Hume writes, “We cannot form to ourselves a just idea of the taste of a pine apple, without having actually tasted it.” (T1.1.1) He also writes that we have no “just

---

22 This is famously disputed by Kant in the Critique of Pure Reason.
idea” of efficacy. (T1.3.14) Again, Hume holds that “we can never have reason to believe that any object exists, of which we cannot form an idea.” (T1.3.14, p. 222) Must the idea be a “just idea” or could we have reason to believe that an object exists of which we have only an “unjust” idea? It seems we can certainly have reason to believe that the taste of a pineapple exists without having tasted it. Similarly, Locke writes, “A blind Man may... sort Things by their Colours.” (Locke, 445) We cannot have a “just idea” of the taste of a pineapple, and the blind man cannot have a “just idea”, or an image, of red. Yet he has an idea of it in some sense, which is sufficient to give him a reason to believe that it exists. By parity of reasoning, it seems we can also have reason to believe that probabilities or powers exist without having a just idea of them, or having ever had an impression of them. Perhaps our ideas of these things are acquired indirectly, like the blind man’s idea of ‘red’.

The blind man’s idea of ‘red’ can be traced to his impressions upon hearing other people use the term “red,” so its meaning for the blind man is indirectly derived from the ideas that other people have of it. Hume allows that we have ideas of ideas: “We can form secondary ideas, which are images of the primary,” whether ideas of our own ideas or of the ideas of others. (T1.1.1) In contrast, for Hume, no one has ever had an impression of causal power or probability on which our idea of it might piggyback. But just as the blind man’s idea of ‘red’ defers to facts about the world with which he is unacquainted, so the meaning of ‘power’ and ‘probability’ may reach beyond the inspectable features of the mental image at hand. It seems that we can refer to “probability” and “necessary connection” and have reason to believe that they exist without anyone having defining images of them, perhaps based on analogy.
We do not always need to give reductive definitions of expressions like “probability” in terms of particular perceptions. Hume may be right that the sentiment of “a customary connexion in the thought or imagination between one object and its usual attendant” is “the original of” the idea of necessary connection, but this does not imply that this customary connection in thought is what necessary connection is. (Hume, E7.2)

Moreover, just as we have unjust ideas of impressions that are absent to us (like the flavor of fruits we have never tasted), that we understand as analogous to impressions we have actually had, we may have an unjust idea of necessary connection, that we understand as analogous to “explanation” in mathematics. Hume writes in the voice of Philo:

It is observed by arithmeticians, that the products of 9, compose always either 9, or some lesser product of 9, if you add together all the characters... To a superficial observer, so wonderful a regularity may be admired as the effect either of chance or design: but a skillful algebraist immediately concludes it to be the work of necessity... Is it not probable, I ask, that the whole economy of the universe is conducted by a like necessity?²³ (Hume 1993, 73)

Just as we give mathematical explanations of numerical regularities, and often assume that they exist even when we do not know what they are, we understand powers as an explanation of a certain kind, though we do not know precisely how they explain the constant conjunction of causes and effects. Some things are the same over time for reasons that can be understood a priori. This constancy is not surprising, and is expected by an ideal reasoner. Thus, perhaps, part of the origin of our notions of causal power and necessary connection is our experience of grasping mathematical

²³ Philo rejects this reasoning, mentioning it in the course of a reductio ad absurdum, but his only argument against it is that it is only found appealing to people of a “metaphysical head.”
and other *a priori* explanations.

A non-circular solution must appeal to something that is likely to continue to persist for some time, given that it has long persisted up to that time.\(^{24}\) This thing has not been directly observed, and there is no automatic given that it is even observable (on any of the various levels of observability).\(^{25}\) Without allowing appeal to something not directly observed, such as an abstract idea of ‘necessary connection,’ ‘power,’ ‘probability,’ or the like, a solution to the problem of induction would be much more difficult and perhaps impossible. But it seems that we have an “innate seed” of a concept of probability, which allows us to make sense of statements of probability in mathematics and everyday life. Moreover, the question of the meaning/meaningfulness of ‘probability’ is not unique to problems of induction. (See Ch. 2)

Hume’s disbelief in the possibility of solving the problem of induction stems from his radical empiricism, which led him to disregard the intuitive meanings of concepts like ‘power,’ ‘connection,’ ‘probability,’ etc., replacing them with concepts that fit more neatly into his theory of abstract ideas. But Hume’s theory of ideas is too restrictive, unable to capture the complexities of the ideas we really have.

1.1.4. Hume’s Fourth Argument for Inductive Skepticism

Hume writes, “If, after reflection, you produce an intricate or profound argument, you, in a manner, give up the question, and confess, that it is not reasoning which engages us to suppose the past resembling the future.” (E4.2.33, p. 39) Knowledge of a

\(^{24}\) I deliberately leave the exact length of persistence needed vague.

\(^{25}\) It might be like a rule of arithmetic, not obviously directly observable even in principle, because of its transcendence/standing outside of the relation of time (and other relations differentiating observed from unobserved). If an equation holds true, such 4 + 3 = 7, manifested in any particular arrangement of objects, it holds true for substitutions of different objects at all times.
reasoned solution to the problem of induction seems to play no role in people’s everyday inductive practices, making the suggestion that there is a solution to be found seem superfluous. Hume expands on this, writing, “I cannot now discover an argument, which, it seems, was perfectly familiar to me, long before I was out of my cradle.” (E4.2.33, p. 39) Animals, too, reason inductively, according to Hume, but we do not take them to be aware of an argument justifying induction:

For if there be in reality any arguments of this nature, they surely lie too abstruse for the observation of such imperfect understandings; since it may well employ the utmost care and attention of a philosophic genius to discover and observe them. Animals, therefore are not guided in these inferences by reasoning; neither are children; neither are the generality of mankind, in their ordinary actions and conclusions: neither are philosophers themselves, who, in all the active parts of life, are, in the main, the same with the vulgar, and are governed by the same maxims. (Hume, E9)

In relying on induction, an adult philosopher is apparently just continuing her lifelong, ingrained habit. She does not typically solve the problem of induction at a certain age and begin acting exactly the same way with a completely different motive, viz. reason instead of instinct/habit.

This argument contributes to the problem of induction indirectly. Custom alone does not bestow normativity, without which there is no rationality or epistemic probability. Hume, in saying that the use of induction relies on custom, implies that the existence of a non-circular argument for induction is not needed to explain our reliance on it. There seems to be no explanatory role remaining for the supposed normative demand to follow inductive reasoning because its predictions are likely to come true. If this normative dimension is not needed in such an explanatory role, one might naturally conclude that we should abandon it. But this would lead to absurdity; as Hume recognized (outside of his radically skeptical stance), some predictions are
clearly more rational than others.

It seems that a person can rationally use induction in practical and theoretical reasoning without knowing a solution to the problem of induction. But how could this be? Intuitively, we feel confident that we genuinely understand and know that the Sun will probably rise tomorrow; our success in inductive reasoning is not due to tremendous luck in blind guessing. But how can this count as understanding, or anything other than luck, if we do not know a solution to the problem of induction? The worry is that even if a consensus were achieved on a solution to the problem of induction among philosophers, it is not clear how this would justify the use of induction for everyone else. It seems we will have to back away, at least to some extent, from the thought that the justification of our beliefs must lie in our knowledge of the grounds of their justification - without giving up on the idea that justifying grounds exist.

Alternatively, the person trying to solve the problem of induction might maintain that her argument is already known universally. Keynes bites the bullet: “We need not lay aside the belief that this conviction gets its invincible certainty from some valid principle darkly present to our minds, even though it still eludes the peering eyes of philosophy.” (Keynes, 264) Like Keynes, I am agnostic about whether people are generally aware of an argument justifying induction, which they are unable to articulate to the skeptic’s satisfaction. However, it would be best if the solution to the problem of induction were simple, making it seem at least somewhat plausible that people are already aware of the solution, on some level, before hearing it articulated. D.C. Williams says that “the solution of the problem of induction must be at bottom as
banal and monolithic as the process of induction itself.” (D.C. Williams, 21) This may be a slight exaggeration, but at best the solution would be easily understandable and perhaps feel familiar or obvious.

Hume seems to assume that a person needs to know the rational justifications of her beliefs in order for her beliefs to be rationally justified. This assumption is plausible but not indisputable. Perhaps what is required for rational belief is less than knowledge of all of the rational principles warranting one’s beliefs. Reliabilists hold that a person can be rational and have knowledge, if his/her beliefs are reliable (probable), even if he/she does not know why they are reliable. For example, beliefs produced by a mechanism somehow evolutionarily “designed” to produce beliefs that are likely to be true might suffice.

This can be qualified. Perhaps inductive knowledge requires both that induction is inherently reliable, and that this reliability is causally connected, in the right way\(^{26}\), with our belief that it is reliable.\(^ {27}\) Why induction is inherently reliable, if it is (and as it seems to be), is a different question.

1.1.5. Hume’s Fifth Argument for Inductive Skepticism

Finally, Hume offers a challenge to induction based on diminishing higher-order probabilities. Unlike #2, 3, and 4, this is entirely distinct from Hume’s main problem of induction (#1). Hume writes:

\(^{26}\) A belief in a fact can be caused by the fact without being justified. John Perry provided the following example in comments on an earlier draft: A package for you is put in the mailbox next to yours by mistake, where it is the only package. You see the package, and mistake the mailbox it is in for your own. Thus you believe that you have a package. If you hadn’t had a package, you wouldn’t have seen a package in the box and thus wouldn’t believe that you have a package. The belief that you have a package is caused by the fact that you have a package, but does not seem to be justified.

\(^{27}\) We can imagine a world in which induction is reliable, but our belief in induction is based upon a misunderstanding and is independent of induction’s reliability. In this case, we might not have inductive knowledge of probabilities or rational beliefs about probabilities, though our beliefs about probabilities would be reliably true.
Having thus found in every probability, beside the original uncertainty inherent in the subject, a new uncertainty deriv’d from the weakness of that faculty, which judges, and having adjusted these two together, we are oblig’d by our reason to add a new doubt deriv’d from the possibility of error in the estimation we make of the truth and fidelity of our faculties. (Hume, T1.4.1, 233)

First we estimate the probability of an event. Then we recognize a degree of uncertainty about our probability estimate, arriving at a higher-order probability, which must be multiplied by the original first-order probability to arrive at our new first-order probability. Then “we are oblig’d by reason” to add a third-order doubt, resulting in a probability of a probability of a probability, eventually diminishing to nothing. Hume’s argument here closely resembles an argument made by Sextus Empiricus:

Those who claim for themselves to judge the truth are bound to possess a criterion of truth... whence comes it that it is truthworthy?... And, if it has been approved, that which approves it, in turn, either has been approved or has not been approved, and so on.” (Sextus Empiricus 1957, I.340, p. 179)

In each case, we recursively challenge the standards upon which the next lower-order judgment of truth or probability is based. Once we raise the question of what reason we have to accept our own judgments, we begin an unending downward spiral. After we assign a probability to an event, we must then assign a second-order probability to that first-order probability assignment being correct. And so on down the spiral, until our initial probability is completely diminished (bracketing the possibility that the subsequent iterations of probabilities approach an asymptote).

This “judgment of judgment” skeptical argument is prima facie distinct from the problem of induction’s multi-level higher-order charges of assuming similarity across the mode of difference between inductive premise and conclusion (i.e. the problem of induction). Both involve iteration, but this is as far as the similarity goes.
In the judgment of judgment case, probabilities are established and then progressively decreased, whereas in the problem of induction, as soon as a probability is suggested it is immediately annihilated by the charge of circularity in assuming similarity between observed and unobserved.

1.2 Other Forms of Skepticism about Induction

1.2.1 Deductivism

Another problem that should be distinguished from the problem of induction is the challenge from deductivism. Many philosophers have taken Hume to hold that any process of reasoning, including induction, could only be rationally justified by deduction, perhaps projecting their own opinion onto Hume. Graciela de Pierris criticizes the view held by David Stove and others that “Hume is a deductivist who condemns inductive inferences because they fail to necessitate their conclusions.” (De Pierris, 2002, 517) Max Black seems to make the same mistake: “The simple argument that Hume desires is a deductive argument.” (Black 1954, 160) To the contrary, De Pierris argues convincingly that “avoiding circularity in our argumentation is not a demand exclusively upheld by the rationalist/deductivist.” (De Pierris 2002, 520) Hume, following Newton, adopted a principle of the uniformity of nature according to which the fundamental laws are exceptionless. But at the same time, as long as our information is incomplete, we can never be sure that we have discovered these ultimate, exceptionless laws (rather than, for instance, mere approximations that hold only in the limited contexts we have observed so far). Russell explains, “Even if some law which has no exceptions applies to our case, we
can never, in practice, be sure that we have discovered that law and not one to which there are exceptions.” (Russell, 1997) For Hume, the particular exceptionless laws that we accept “are always in principle revisable,” as new data is acquired. (De Pierris 2002, 540)

Therefore, even if we could justify Hume and Newton’s principle of the uniformity of nature, which Hume holds sufficient for solving the problem of induction, inductive inferences would still not strictly entail their conclusions. Therefore, it should not be assumed from the outset that an inductive inference can only be justified by argument if it can be shown to be deductively valid. Moreover, the fact that an argument is non-deductive does not necessarily render it any less objectively correct.28

1.2.2 Skepticism in Philodemus and Goodman

Hume’s problem is not a reference class problem like the puzzles of Nelson Goodman and the ancient Stoics. These problems concern how to choose among or balance different classificatory schemes in forming an inductive premise, given that different classifications may yield incompatible conclusions. For example, Philodemus writes, “It is unclear to us whether a thing is this *qua* this.” (Philodemus, 98) This can be interpreted in at least two ways. First, in reality, unlike Groundhog Day, the world does not appear today *exactly* as it did yesterday. Perhaps the Sun has risen in my experience only because (*qua* the fact that) Richard Dawkins has always been under the age of 70 (he turns 70 tomorrow). This fact seems irrelevant, but perhaps it is not.

28 Franklin helps to put this to rest, pointing out that “logical relations weaker than entailment” hold even “between the necessary truths of mathematics.” (Franklin, 1987) For example, the Riemann hypothesis is likely to be correct, relative to the current state of mathematics. See also Keynes 2009, Mackie 1979, and Polya 1954 for similar arguments.
Perhaps, as a result of this seemingly irrelevant fact, the Sun will not rise tomorrow. Thus we can interpret Philodemus’s “qua” skepticism as expressing the worry that some difference between the observed and unobserved that we do not include in our selective inductive premise may turn out to be relevant.

Unlike the problem of induction, this skepticism dissolves upon the ideal inductive conditions of Groundhog Day, suggesting that the problems are distinct. Perhaps some difference between my psychological structures today and yesterday creates different probabilities for our respective tomorrows. As long as there are differences in the epistemic givens, this doubt can be raised. But in Groundhog Day, there are no differences in our epistemic state that could turn out to be relevant. Yet the problem of induction can still be raised; we can still ask why I should believe that tomorrow is likely to resemble today. Groundhog Day is a more sterile environment in which to raise the problem of induction, because it is immunized against these externalities.

Alternatively, we might interpret the statement that “it is unclear to us whether a thing is this *qua* this” as a form of the so-called “modal objection”. Perhaps the Sun has risen regularly in the past *qua* past, or “qua” some other modal restriction, such as being observed. Unlike the above interpretation, the modal objection applies to Groundhog Day. However, in Groundhog Day, the difference between past and future or any other reflexively represented fact cannot influence the epistemic probability, as this would contradict the principle that epistemic probability supervenes on psychological structure.
The Stoics foreshadowed Goodman’s new riddle, according to which any hypothesis about the unobserved can be inductively confirmed by any observations depending on how the classes are demarcated. “The Stoics ask on what grounds inferences between some classes of things are preferred to inferences between other classes.” (De Lacy, in Philodemus, 158) Similarly, Goodman writes, “Merely to belong to some one class is not enough; for any two things belong to some one class.” (Goodman, 44) Epicureans reply that “the inference must be between classes that are as similar as possible.” (De Lacey, in Philodemus, 158) Hume follows Epicureans on this point. Goodman asks what it means to be similar: “For just when are two things of the same kind?” (Goodman, 44) Goodman takes the reference class problem to its logical conclusion with the example of “grue,” a provocatively counterintuitive way of demarcating a reference class.

In response, we use the reference class ‘green’ in our inductive inferences rather than ‘grue’ because we believe that green is a natural kind and grue is not. Goodman asks us to defend this judgment; but it is not clear that a defense is needed. Many philosophers, including Hume, hold that we recognize resemblance intuitively and immediately. Locke writes, “A Man infallibly knows, as soon as ever he has them in his Mind that the Ideas he calls White and Round, are the very Ideas they are, and that they are not other Ideas which he calls Red or Square.” (Locke, 526) It seems that we “can just tell” that a green impression and a later blue impression are not similar experiences in the way in which two green impressions are similar29, despite the possibility of labeling both with the same term, “grue”.

29 However, they can be regarded as being similar to a certain degree: “BLUE and GREEN are different simple ideas, but are more resembling than BLUE and SCARLET.” (Hume, T1.1.7)
Quine and Mackie share this view. Quine writes, “Two green emeralds are more similar than two grue ones would be if only one were green.” (Quine, 6)

Similarly, Mackie writes:

We can stick to the common-sense view that things which are green (especially if they are of the same shade of green) at different times resemble one another in a way that things called grue by Goodman’s definition at different times may not. And then it is plausible to suppose any formal principles of probabilification there may be will take account of the presence or absence of real resemblances rather than the merely syntactical forms which ‘grue’ can satisfy as well as ‘green’. (Mackie 1979, 167)

Goodman’s riddle reveals that resemblance cannot be captured purely “syntactically,” because syntax alone cannot distinguish terms like “grue” from terms like “green”. But the riddle dissolves if we allow that there is more to meaning than syntax, and agree with Hume that we intuitively recognize resemblances between impressions.

Evan Fales criticizes the view that Goodman’s new riddle of induction involves “a verbal trick – predicates like grue fail to capture genuine similarities,” suggesting that someone holding this position is committed “to a realist position with regard to universals.” (Fales, 2004) But the notion of similarity need not be analyzed in terms of universals. It could be analyzed in terms of tropes, or properties and substances, or natural kinds, or some other alternative. Or, similarity might be unanalyzable.

Even if the reader is unwilling to concede that Goodman’s “new riddle” presents no genuine puzzle because it involves a mere verbal trick, it should be agreed on all sides that Goodman’s riddle is distinct from Hume’s problem of induction, in that it is not the puzzle that Hume intended. Hume, of course, took resemblance as primitive, so it is inappropriate to object to a proposed solution to Hume’s problem of induction on the grounds that it does not answer Goodman’s challenge to the meaning
of resemblance. In responding to a purported solution to Hume’s problem, it is beside the point to raise Goodman’s riddle.

Goodman’s questioning of the principle behind the distinction between “green” and “grue” implicitly questions Hume’s assumption that people recognize resemblance intuitively. De Pierris explains that in Hume, “the relation of resemblance between two ideas or two impressions... is classified as an a priori relation.” (De Pierris, 2002, 514) She explains Hume’s account of how we come to know about resemblance: “In one single act of the mind, we ‘observe’ internal features of items before the mind, compare them, and immediately and directly establish the relation in question.” (De Pierris 2002, 508) The idea of regularity depends on the idea of resemblance, so unless we allow Hume’s assumption that we intuitively recognize resemblances, the inductive premise of vast regularity in our past experience cannot be given, and the problem of induction cannot even be raised. Thus Goodman’s new riddle is more radical than Hume’s problem in this respect. Still, many philosophers imply without argument that the problems are the same. For example, William Edward Morris writes that Goodman’s riddle “can be regarded as a particularly dramatic way of posing Hume’s problem,” without further explication. (Morris, 459) But I would argue that Goodman’s riddle is a dramatic way of posing a different problem.

Admittedly, there is an appearance of similarity between “gruesome” predicates and time-restricted laws (such as a law of gravitation that ends at a certain time), which might seem to suggest that Hume’s problem and Goodman’s riddle are the same, given that the time-restricted law objection (according to which past regularities might be explained by a time-restricted law instead of a universal or time-
impartial one) applies to attempts to solve Hume’s problem.\footnote{Alistair Isaac pointed out this apparent similarity to me. For more on the time-restricted law objection, see the discussion of John Foster’s account in Ch. 3.} But gruesome predicates and time-restricted laws are different. “Grue” refers to different colors at different times, and time-restricted laws prescribe different regularities at different times. In one case the change is linguistic, while in the other it is empirical. A world with time-relative laws, in which everything green instantly turns blue in the year 2100, is not a world in which ‘green’ ceases to be a simpler idea than ‘grue’, even if ‘grue’ turns out to be more scientifically useful because there are more grue correlations than green correlations. Commonsense holds that all green objects, but not all grue objects, are the same color, even in a world with shifting (time-restricted) physical laws. One can accept the time-restricted law objection without accepting Goodman’s riddle. Someone who accepts primitive universals, tropes, or resemblance relations still faces the possibility of time-restricted laws which cut across them. Adopting primitive universal properties is different from adopting universal laws.\footnote{This holds even for Armstrong, for whom laws are connections between universals. There could, conceivably, be universals without such connections.}

1.2.3 Hempel’s Paradox of the Raven

The statement that all ravens are black is logically equivalent to the statement that everything not black is not a raven. A green leaf is not black and not a raven, and so if we accept universal inductive inference (as opposed to singular inductive inference), the observation of a green leaf confirms the universal generalization that everything that isn’t black isn’t a raven, and so, paradoxically, confirms the equivalent statement that all ravens are black.

I once thought this paradox could be solved by considering that in this example
we presumably do not know the number of objects in the world. If we suppose that we know that the world contains only 100 (or any finite number of) objects, then the hypothesis that all ravens are black is somewhat confirmed by the observation of a green leaf, by “linear attrition”, i.e. the elimination of possible counterinstances. McGrew (2001) dubs “the thesis of linear attrition” the fallacious solution to the problem of induction according to which a universal is confirmed by the observed cases because there are fewer cases left to check in which the universal might be falsified. If we know that the world contains 100 objects, the observation of a green leaf makes it slightly more likely that all ravens are black because there is one fewer object that might be a non-black raven.

Dretske explains why the thesis of linear attrition does not solve Hume’s problem of induction: “Confirmation is not simply raising the probability that a hypothesis is true: it is raising the probability that the unexamined cases resemble (in the relevant respects) the examined cases.” (Dretske, p. 256) But though it does not help us solve Hume’s problem of induction, the fact that there are only 99 possible counterexamples to the hypothesis that all ravens are black instead of 100 does, slightly, increase the probability that all ravens are black. Thus, when we know the total (finite) number of objects, the observation of a green leaf slightly confirms the proposition that all ravens are black.

On the other hand, if the total number of objects is unknown but presumed finite, then the observation of a green leaf gives us, in addition to the information that there is one fewer possible counterexample to our hypothesis, the information that there is one more object than we previously knew about. Thus there is no attrition in

---

32 See also Foster, p. 25, and Goodman, p. 69.
this case; the expected number of remaining (unobserved) objects in the universe is no less than before the green leaf was observed. I used to think that this could explain how the observation of a green leaf does not increase the probability that all ravens are black even slightly, in the case in which the total number of objects is only known to be finite. However, the mystery remains for the case in which the number of objects in the world is known to be infinite. Intuitively, in this case as in the others, observing a green leaf would not increase the probability that all ravens are black. However, in this case, observing a green leaf would not increase the number of objects known to exist, so it remains mysterious how this observation could fail to confirm the hypothesis that all ravens are black.

I have not seen a solution to Hempel’s problem that I find satisfying. Like Goodman’s riddle, Hempel’s puzzle is problematic because it suggests that accepting ordinary inductive standards leads to the absurd conclusion that everything confirms everything else (or nearly so). An observation of a green leaf confirms not only the generalization that all ravens are black, but also the generalization that all ravens are purple.

However, it seems that Hempel’s paradox of the ravens is not Hume’s problem of induction. *Prima facie*, the two puzzles have little in common, except perhaps their shared skeptical conclusion. Moreover, Hempel’s paradox seems to arise only for induction involving universal generalization. To show that the Sun will probably rise tomorrow is to solve Hume’s problem of induction. But establishing the truth of this singular proposition does not require accepting universal generalization.
1.2.4 Cartesian Skepticism

“The principle reason for doubt,” says Descartes, is “my inability to distinguish between being asleep and being awake.” (Descartes, 61) This is distinct from Hume’s problem of induction, which allows us to assume that our memories are correct, as part of our inductive premise. Moreover, the same rational standards apply to induction whether we are asleep or awake. In dreams, our inductive premises tend to be weaker: our memories are less clear and less consistent, and regularities are less stable. But to the extent that our memories are accurate, clear and stable while dreaming, the distinction between dreaming and being awake is not relevant to the problem of induction. For example, the difference between a dreamed Groundhog Day and a real Groundhog Day is irrelevant to the rationality of induction. If induction is rational in one, it is rational in the other.

Descartes imagines that “some malicious demon of the utmost power and cunning has employed all his energies in order to deceive me.” (Descartes, 15) This enables him to discover what can be known beyond a doubt. Hume, in contrast, from the point of view from which he states his problem of induction, does not doubt everything that can be doubted but takes certain things for granted. For example, Hume takes memory, present impressions, and intuitions of resemblance for granted. We do not need to worry about whether the inductive premise is true, in this context. We can assume that our observations and memories are accurate depictions of extensive regularities (at least among our ideas) and still raise Hume’s problem of induction.

Moreover, given the assumption that our memories of impressions are
accurate, an evil demon could not fool the inductive reasoner very much in the long run, since the demon would need to set up the eventual deception with a large number of instances in which the inductively likely thing occurs as predicted, in order to make the inductive reasoner very surprised when regularity is finally cut off. The more regular our world has been, the more surprised we will be when the evil demon pulls the rug out from under us. But if the world has been very regular, our inductive inferences will have been successful many times before our eventual upset. Not even an all-powerful evil demon could cause a consistent inductive reasoner with good memory to be surprised most of the time.\(^{33}\)

1.3 The Paradox of Induction

Everyone believes that the Sun will probably rise again, based on its past regularity (and regularities in the world in general). Surely this counts as knowledge. But any argument that the future probably resembles the past, on the grounds of this past regularity, must circularly presuppose a (higher-order) resemblance of the past to the future, reasoned Hume. For example, it must assume that the future resembles the past in the sense of resembling its own relative past (Russell) or in the sense of being accompanied by secret powers (Hume). Thus our inductive probability knowledge seems groundless, based on no good reasoning.

Hume intends this as a paradox. Despite his seemingly compelling radically skeptical arguments against induction, Hume believes that the Sun will rise again,

\(^{33}\) There are two possible cases. First, the demon cuts off regularity immediately after the first inductive inference is made, but after there has been much regularity in the world. This leads to a reference class problem that is distinct from Hume’s problem of induction and ruled out by Groundhog Day (See pp. 33-34 of this dissertation). Second, the demon cuts off regularity only after many successful inductions have been made. This is the case being considered here.
endorsing “in no uncertain terms the normative use of causal reasoning.” (De Pierris 2001, 351) Hume’s “sceptical solution” does not, and is not intended to, resolve the paradox. To give a merely naturalistic, e.g. psychological-habitual, or evolutionary, explanation of our confidence in induction would not only explain but seem to “explain away” our inductive judgments, failing to justify them. “It is no justification of inductive belief to show that it is a habit.” (Reichenbach 1938, p. 347) In addition to the existence of our inductive judgments, their success needs to be explained. Surely their past success is not just a coincidence.34

Moreover, in addition to the existence and past success of our inductive inferences, their rationality needs to be explained: we not only do form probability judgments based on inductive means, and do so successfully (so far), but, commonsense holds with great confidence, we should. We all believe that our judgment that the Sun will probably rise tomorrow is correct.35 But if the belief that the Sun will probably rise tomorrow could be “explained away” without reference to its objective correctness, such that we would have this belief whether it were correct or not, this would undermine the belief that it is correct, which is tantamount to undermining the belief that it is true. At the same time, if our inductive judgments were not correct, i.e. not epistemically likely to be successful, there could be no coincidence-lessening explanation of why they have been so successful.

1.3.1 Pyrrhonian Skepticism

34 The fact that we make inductive judgments can perhaps be explained by evolution, given the assumption of enough regularity in the past for natural selection to have taken place. But their success, in terms of the fact that the regularity continues as predicted, is another story.

35 To believe “x, but it is not epistemically correct to believe x” effectively raises Moore’s paradox.
Hume perceives induction through the incompatible lenses of radical and mitigated skepticism. He adjusts his skeptical position by accepting different assumptions at different times. His radical skepticism is purely theoretical. Hume does not endorse the Pyrrhonian skeptic’s position that radical skepticism about induction should be applied to practical affairs: “None but a fool or madman will ever pretend to... reject that great guide of human life.” (E4.2.31, pg. 36.) Pyrrhonian skeptics endorse the use of radical skepticism in practical life, saying it engenders tranquility. (Sextus Empiricus 1961, Outlines of Pyrrhonism, I, Ch. 4, paragraph 9, p. 7) But Hume disagreed, calling Pyrrhonism “excessive” and saying it “vanishes like smoke” as soon as met “by the presence of the real objects,” adding that the skeptic should be restrained to “his proper sphere.” (E12.2.126-7, pp. 158-159.)

In his voice of “mitigated skepticism,” representing his practical and scientific beliefs, Hume follows Newton. (See De Pierris 2002, 521.) Newton’s rule 3 states:

Rule 3: The qualities of bodies, which admit neither intensification nor remission of degrees, and which are found to belong to all bodies within the reach of our experiments, are to be esteemed the universal qualities of all bodies whatsoever. (Newton, Book III, p. 384)

Newton adds that we should not “recede from the analogy of Nature,” which is “always consonant to itself.” (Newton, p. 384) Although Hume the radical skeptic taunts that no reasoning a priori or a posteriori can justify a belief in the uniformity of nature beyond what we have already experienced, Hume the mitigated skeptic agrees with Newton’s principle of uniformity that nature is “always consonant to itself.” But he denies that there is any justification or rational basis for confining radical skepticism to theoretical inquiry, though it is undoubtedly good to do so.
The paradox is that there can be no non-circular argument for this endorsement of induction within the mitigated skeptical stance, so reason seems to pull apart from reality, failing to provide us with guidance about what we should believe. Hume writes in the Abstract, “Philosophy would render us Pyrrhonian, were not nature too strong for it.” (Hume, Abstract of the Treatise, paragraph 27) He writes that the step needed to render any argument for induction non-circular could never be produced: “However easy this step may seem, reason would never, to all eternity, be able to make it.” (Hume, Abstract of the Treatise, paragraph 16) Writing as if there is an unbridgeable gap in the reasoning that grounds our trust in induction, Hume leaves us without any palatable options, the hallmark of a paradox.

1.3.2. Popper, Strawson, Reichenbach

There are many approaches to Hume’s problem, including solution by non-circular justification (e.g. J.L. Mackie and Laurence Bonjour), rejection of the assumption that induction is justified/worthy of endorsement (e.g. Karl Popper and the ancient Pyrrhonian skeptics), and lowering the bar for “justification” (e.g. Hans Reichenbach and P.F. Strawson). Only the first line of response holds promise for resolving the paradox, I shall argue.

Popper writes that “The belief that we use induction is... a kind of optical illusion.” (Popper, 104) He argues that instead of induction we use trial and error, subjecting theories to “the severest tests” in order to try to falsify them, and then accepting the theory that remains standing. (Popper, 104, 102) Popper’s view that science and commonsense do not rely heavily on induction is implausible. Popper’s supposed alternative to induction, falsification, tacitly assumes that unobserved
instances will not conform to theories that have been falsified with respect to past instances.\textsuperscript{36} From a counter-instance, we can deduce that a theory does not hold true universally. But we cannot deduce that the theory does not hold true for every unobserved instance. How can falsification tests in the past tell us anything about the future? On what grounds can we rule out theories holding true with respect to the future merely because they have been counter-instanced in the past? Only if we assume that there is some theory which will never be falsified, which it is the goal of science to discover or approach, can there be any reason to reject with respect to the future theories falsified in the past. It seems that Popper, ironically, tacitly relies on an assumption of the uniformity of nature that is subject to Hume’s circularity objection.

Even if we grant that Popper’s falsificationist account correctly describes scientific methodology, there is no reason to favor science over pseudoscience unless we presuppose that theories falsified by an observation are less reliable predictors of the yet-to-be observed than theories yet to be falsified.\textsuperscript{37} Without this presupposition, the predictive failure of astrology in our experience would be entirely irrelevant to the question of whether it is likely to yield predictive success in the future. Without this

\textsuperscript{36} For a similar argument, see Foster, pp. 10-11.

\textsuperscript{37} Thomas Ryckman objected to an earlier version of this argument along the following lines: “Popper treats laws rather than theories, laws considered as universal generalizations. Hence, by modus tollens, one counter instance serves to negate the law. The observation of a single white raven falsifies the ‘law’ that all ravens are black. There is no need or reason to worry about further instances either unobserved or observed.” In response, let us distinguish between laws, non-laws, and law-candidates. Once a law-candidate is falsified, it becomes a non-law. Laws cannot be falsified, because the world contains no counter-instance (by the meaning of “law”). The question is why we should think that law-candidates that have shown themselves to be resistant to falsification in past experiments provide a better guide to the future than falsified non-laws. (John Perry helped formulate this response.)

Suppose we have two hypotheses about a universal law, A and B. If we find a counter-instance to A in the past, but no counter-instance to B in the past, we know that A is not a universal law, but B might be. Still, perhaps the future will occur as if A were a universal law - perhaps the observed exception is the only one. How do we know that this is less likely than that B is a universal law? A counter-instance negates a law, but cannot give us information about what is likely to happen in unobserved instances, unless we question-beggingly assume that universal laws are likely to exist.
tacit principle of uniformity, consistently falsified theories are on no weaker ground with respect to unobserved instances than theories that have proven resilient against falsification tests, and a falsificationist science would be fruitless, giving us no basis for prediction or manipulation. Popper’s view that scientific inquiry proceeds through falsification of competing hypotheses may have been largely correct, but can tell us nothing about Hume’s justificatory challenge at T1.3.6 and E4.2.

P.F. Strawson, in contrast to Popper and the Pyrrhonian, holds that induction is justified. But he evades the problem of induction by maintaining that the rationality of induction is known analytically and immediately, requiring no further demonstrative or experiential argument. Even if we accept that there is no actual or possible non-circular argument grounding induction, it does not immediately follow that induction is not justified. Some things, perhaps, are justified without argument. For Strawson, the rationality of induction is known as a matter of meaning, like the rationality of deduction, so there is no need for it to be grounded in further argument. He writes that induction is rational as “a matter of what we mean by the word ‘rational’ in its application to any procedure for forming opinions about what lies outside our observations... to have good reasons for any such opinion is to have good inductive support for it.” (Strawson 1963, 262) He says it is “an analytic proposition... that, other things being equal, the evidence for a generalization is strong in proportion as the number of favourable instances, and the variety of circumstances in which they have been found, is great.” (Strawson 1963, 257) On this view, the pre-argument default is that induction is justified, its rationality assumable. Analogously, were an argument needed for the justification of deduction, nothing less than an infinite regress
of arguments would suffice; for example, nothing less than an infinite regress of
question-begging applications of modus ponens could justify modus ponens. But we
give deduction the benefit of the doubt, allowing that it is justified without argument.
Why not take the same position with respect to induction? For Strawson, induction
and deduction are equally primitively justified.\footnote{This is not to be confused with
the view commonly held by advocates of the logical theory of probability that probabilities are lesser degrees of, and ontologically at least as basic as, logical entailment. (See Chapter 2.) There is a difference between thinking induction is justified primitively and thinking that the idea of probability is primitive.}

Hume was not in agreement with Strawson. For Hume claims that an
inductive inference is not a “tautology”. (E4.2.32, p. 37) According to Hume, the idea
of the cause does not contain the idea of the effect; they are separable ideas. If it were
ture by definition that the future is likely to resemble the past, this would imply that
“the meaning of a statement about the future is a statement about the past.”
(Reichenbach 1938, 343) Strawson avoids this absurd conclusion by locating the
analyticity not in the semantic containment of the inductive conclusion in the
inductive premise, but in the meaning of the terms “rational” and “reasonable”. He
writes that forming beliefs based on inductive reasoning is “what ‘being reasonable’
means in such a context.” (Strawson 1963, 257)

This puts the cart before the horse. It seems that induction is rational not as a
matter of the meaning of “rational”, as Strawson contends, but rather because
induction is likely to yield successful predictions. It is reasonable to believe 4+4=8
because it is true; it is not the case that 4+4=8 is true because it is reasonable to
believe it. Analogously, it is reasonable to believe that the Sun will rise because it is
probable. The reasonableness of accepting conclusions based on induction is not built
into the meaning of “reasonable”; rather, the reasonableness of accepting conclusions that are probable is built into the meaning of “probable”\(^\text{39}\). The tautology lies in the relation between probability and reasonableness, not induction and reasonableness. Thus the challenge is to show that inductive conclusions are in fact probable.

Strawson understands the problem of induction as the need to show that using induction is rational, justified, or reasonable, holding that the rationality of induction is tautological. But we can just as well pose the problem in terms of the need to show that the Sun is likely to rise tomorrow, i.e. in terms of probability. Yet it does not seem to be a tautology that the Sun will probably rise tomorrow because it has risen regularly in the past-- it does not seem to be a matter of the meaning of “The Sun has risen regularly in the past” that it will probably rise again. (See Hume, E4.2.32, p. 37) If beliefs aim, in some sense, at truth, then the rationality of induction could be tautological only if inductive arguments themselves were tautological. But inductive arguments can not be tautological, because a statement about the past is not a statement about the (probable) future, whatever it may imply about the future. (See Reichenbach 1938, p. 343)

Reichenbach also tries to show that induction is rational without first showing that its conclusions are probable, and so his “pragmatic” response to the problem of induction faces the same problem. He begins with the idea that probability is limiting relative frequency. (Reichenbach 1938, 350) He writes, “The aim of induction is to find series of events whose frequency of occurrence converges toward a limit.” (Reichenbach, 1938, 350) Reichenbach claims that if physical processes in the world converge on limiting relative frequencies, then “the use of the rule of induction is a

\(^{39}\) Mark Crimmins helped frame this point.
sufficient instrument to find” them. (Reichenbach 1949, 474)\textsuperscript{40} Induction offers our only chance of success, if we have any chance at all. “Since we do not know a sufficient condition to be used for finding a limit, we shall at least make use of a necessary condition.” (Reichenbach 1949, 475) Reichenbach gives an analogy to show “the logical structure” of his argument: “‘I do not know whether an operation will save the man, but if there is any remedy, it is an operation.’ In such a case, the operation would be justified.” (Reichenbach 1938, 349)\textsuperscript{41} Reichenbach refrains from affirming that the future is likely to resemble the past, or “that there is a certain probability of success,” just as the doctor might refrain from saying that the operation has any chance of success, claiming only that one should adopt it as a working assumption. (Reichenbach 1949, 473) But if we cannot show that the Sun will probably rise tomorrow, or that observed instances are even relevant to the probabilities of unobserved instances, we have not solved the problem. As Black disparagingly remarks, “The only way, to justify induction, we are told, is to show that induction is rational ‘without reference to the truth or probability of its conclusions.’” (Black 1954, 157) Reichenbach admits that he does not try to prove “that the inductive conclusion is true or probable.” (Reichenbach 1949, 479) This contrasts with William James’s pragmatic definition of truth as “the expedient in the way of our thinking.” (James, 222) Reichenbach, rather, thinks that inductive reasoning is expedient even though it gives us neither truth nor probability.

\textsuperscript{40} Reichenbach adds the caveat that “we would never know whether the observed initial section of the sequence were long enough to supply a satisfactory approximation of the frequency.” (Reichenbach 1949, 480) The point seems to be that the use of induction is sufficient to find the limiting relative frequency if there is a limiting relative frequency, and if enough observations have been made to approximate that frequency - but we can never know whether enough observations have been made.

\textsuperscript{41} On an earlier draft, John Perry expressed doubts about this argument with respect to an operation. I am agnostic.
He writes, “We do not perform... an inductive inference with the pretension of obtaining a true statement. What we obtain is a wager; and it is the best wager we can lay because it corresponds to a procedure the applicability of which is the necessary condition of the possibility of predictions.” (Reichenbach 1938, p. 356-7) This may be an interesting result, if true, but it does not vindicate commonsense. Reichenbach says that his method will succeed only if there exists a limit of the frequency, which he says cannot be proven. (Reichenbach 1949, 472) For all we know, even given all of our observations of regularity, says Reichenbach, “The world may be so disorderly that it is impossible for us to construct series with a limit.” (Reichenbach 1938, 350) For Reichenbach, using induction gives us our best guess; but this does not mean it is even remotely likely to be true, or that the inductive premise is known to be probabilistically relevant to the inductive conclusion to any extent. But Reichenbach’s denial that the inductive premise is known to be probabilistically relevant to the inductive conclusion is incompatible with commonsense. Reichenbach is correct that “we do not know whether the world is predictable.” (Reichenbach 1938, p. 351) Indeed, the Sun may not rise tomorrow. But this does not mean we cannot know that the Sun is likely to rise tomorrow.

Reichenbach’s pragmatic justification of induction is similar to Pascal’s Wager, though more refined and slightly closer to the ordinary notion of epistemic justification. Pascal tried to justify belief in God by the argument that we lose more if we erroneously believe that God does not exist, than if we erroneously believe that God does exist.42 This type of justification, like Reichenbach’s wager, cannot give us

---

42 John Perry pointed out in response to an earlier draft that Pascal assumes “that if there is a God, that God will want to be recognized; it might be shy God, who punishes those who believe.”
knowledge of probability, telling us only what we should believe for non-epistemic reasons. However, Reichenbach’s wager has a slightly stronger claim to being a justification of belief as a truth-oriented epistemic practice. In Pascal’s wager, the goal that belief in God is useful in achieving is escaping eternal damnation, not having probable or true beliefs. In Reichenbach’s wager, in contrast, the goal is to achieve true beliefs about objective probabilities. The problem is that Reichenbach does not try to show that it is (epistemically) likely that inductive methods will allow us to reach this goal. Intuitively, the normative demand to use induction is based on the fact that it is good to have probable beliefs. Reichenbach provides normativity (or at least claims to), but does not promise the probability from which this normativity is normally presumed to stem. Reichenbach can only say that it is a practically wise wager that the Sun will rise tomorrow, not that it probably will.

In summary, for Popper and Strawson, like the Pyrrhonian, there is no problem of induction in need of a solution. Unlike Popper and Strawson, the Pyrrhonian is critical of common behavior, holding that induction, though commonly used, is unjustified. Thus the Pyrrhonian has no need to argue that induction is justified, maintaining that it is not. Popper also denies that induction is justified. But, unlike the Pyrrhonian, he is not critical of common behavior, holding that although induction is not justified, it is not common either. Reichenbach, unlike Popper, Strawson and the Pyrrhonian, recognizes that there is a problem of induction. He takes himself to solve the problem, at least partially; Reichenbach does not accept Hume’s claim that there is no argument for induction. However, like Popper, Strawson, and the Pyrrhonian, Reichenbach is unable to support the straightforward, intuitively undeniable claim that
the future is likely to resemble the past, telling us only what we should believe, even if it turns out to be unlikely. Without a defense of the claim that induction is likely to be successful, we are left with the absurd result that our track record of success in using it is purely coincidental, the result of immensely lucky guessing. Moreover, Reichenbach calls his principle that “sequences of events converge toward a limit of the frequency” a “uniformity postulate,” suggesting that his theory does not offer a non-circular solution to Hume’s problem. (Reichenbach 1949, 473)

None of these responses to the problem of induction is sufficient to justify what commonsense is unwilling to abandon. The problem of induction demands a direct solution. At issue is whether statistical reasoning, reasoning from frequencies in some samples to probable frequencies in others, can be non-circularly grounded. Commonsense demands that there must be a solution, because we know, understand, and have solid grounds for believing that the Sun is likely to rise tomorrow. Many philosophers automatically reject any attempt at a direct solution, perhaps based on Hume’s arguments that the problem is unsolvable. But these arguments are not conclusive. Usually, when faced with a paradox, philosophers look for the solution they assume must exist. The problem of induction should be approached no differently.
Chapter 2. Probability, Coincidence and Explanation

2.1 Probability

2.1.1 Introduction

Consider the informal, epistemic idea of probability (i.e. probability relative to states of knowledge/data/information) that we use in our everyday lives. It is common across widely varying ages, languages and cultures. The Greeks expressed it with the term “εἰκος”. The Latin equivalent was “probabilia”. Centuries earlier, in China, Sun Tzu wrote that under certain circumstances, the chances of victory and defeat in battle are equal, apparently comparing chances to weights and balances. (Sun Tzu, 88, 129)

It should not be surprising that the idea of probability is so ancient and widespread, given its centrality in our lives and close relationship with other basic concepts, like possibility and truth. As Cicero famously says in De Natura Deorum, Probability is “the very guide of life.” Without at least the root of a concept of probability, a person could not think ‘maybe’ or ‘probably’, nor reflectively expect something to happen with other than full confidence. Without probability, much common behavior would be inexplicable. If someone decides to go for a walk because he thinks it will not rain, he cannot make sense of bringing an umbrella “just in case” it rains, without having a probability concept of some kind.43

The concept of probability does not presuppose that induction is justified. Not all probabilistic reasoning is inductive reasoning; probability judgments are often made on the basis of non-inductive arguments. We can judge probabilities based on methods other than induction, such as attributions of symmetry. Despite problems

43 A concept of ‘possibility’ alone does not seem to suffice; fire-breathing dragons are also a possibility, but the person in this example chooses to bring an umbrella and not a shield.
with some formulations of the principle of indifference (to be discussed later in this chapter), it seems that there are some legitimate applications of it.

Consider two examples. First, Mackie writes about the reasoning “which yields judgments of epistemic probability about games of chance” that he says is neither deductive nor inductive. “It makes no extrapolation from observed to unobserved cases, and no inference to a deeper explanation. You need never have played a game of this sort before and you need no theory about it.” (Mackie 1979, 165-166) We reason probabilistically about games of chance, not on the basis of statistical studies of past frequencies, but on the basis of an assumption of the equal likelihood of different symmetrical outcomes. Though Hume describes our psychology as being the same, the criteria used to determine probability in each case (symmetry and induction) seem different.

Second, a perhaps clearer example of symmetry without induction is given by 14th century mathematician Nicole Oresme, who writes that it is unlikely that “the number of stars” in the Universe “is a cube,” because “such numbers are much fewer than others.” (Franklin 2002, 142) It would be a coincidence if the number of stars turned out to be a cube. There is an implicit, and relatively innocent, use of the principle of indifference here, in that we assume that the property of being a cube does not make a number more or less likely to number the stars. (The alternative would seem *ad hoc.*) But there is no use of inductive reasoning based on *a posteriori* observation of repetition: we believe that the number of stars is unlikely to be a cube because of our mathematical assessment that there is a decreasing relative frequency.

---

44 Similarly, Sextus Empiricus says the proposition that ‘The number of stars is odd’ is “neither convincing nor unconvincing.” Against the Professors 7.242-6, quoted in Long and Sedley, pp. 238-239.
of cubes as positive integers increase.\textsuperscript{45} Such reasoning is mathematical, not inductive. We do not need to consult past instances in which we have been confronted with large numbers of objects and found them not to be numbered by perfect squares. Most of us have collected no such data, nor is it necessary. Not all probabilistic reasoning is grounded in induction, so a justification of induction that invokes probability is not automatically circular.

\textit{2.1.2 The Definition of “Probability”}

One might ask if there is one, or more than one concept of probability among commonsense discourse, science, math, law, etc. “Probability” has been carved by many distinctions: epistemic/objective, limiting relative frequency/degree of confirmation (Carnap), quantitative/non-quantitative, symmetry/induction/degree of analogy, etc. The question of whether A and B are identical does not always admit of an exact “yes” or “no” answer, free from decision by convention. Probability concepts significantly overlap and are not therefore radically equivocal, like the term “bank”. But where along the scale from univocality to equivocality does ‘probability’ fall? I am sympathetic to Keynes’s view that “In the great majority of cases the term ‘probable’ seems to be used consistently by different persons to describe the same concept.” (Keynes, 8) If there is a non-vague answer to the question of univocality versus equivocality at all, it seems that we are working with one concept with some different features in different contexts, not different concepts with some analogous features.

\textsuperscript{45} In reality, we may need induction in order to know that the number of stars in the universe is large. But we could, in principle, know this from observation alone.
The everyday notion of probability is not essentially quantitative, nor did probability as we know it first arise in mathematical form. Even science often invokes non-numerical forms of probability. “When Darwin, for example, speaks about probability it can hardly be supposed he has some quantitative theory in mind.” (Franklin 2002, 369) Its nature or essence is hard to capture, but its meaning is understood by all. Augustine’s remarks about time apply equally to probability: “I know that I speak these things in time... How, then, know I this, when I know not what time is? Or is it, perchance, that I know not in what wise I may express what I know?” (Augustine, Book 11) Similarly, I know that the Sun will probably rise, so it seems that perhaps I must know what probability is, even if I cannot express this knowledge.

Some philosophers assume that a conceptual analysis of probability would be needed before the concept of probability could be used in a solution to the problem of induction. But we do not need to be able to give a philosophical or mathematical analysis of probability in order to correctly understand and use the term. “We do not need to be able to express a concept in other words in order to understand it.” (Maher, 10) “The ability to secure a subject matter doesn’t require the ability to define it.” (Crimmins, class, 9/30/09) Moreover, since the ‘probability’ concept is independent of concepts about induction, as discussed above, use of this term in a proposed solution to the problem of induction need not be circular. To an overzealous critic, every word of a proposed solution is presumed the source of circularity in the argument, until proven innocent. This leads to a futile search for so-called “real definitions” of primitive, universally understood terms. We can make well-grounded probability judgments without having a comprehensive theory of probability, just as we can make
judgments of arithmetical equality without a theory of arithmetic. Although there is sometimes value in searching for real definitions of the terms we use, use of a term does not necessarily bring with it the burden of giving the term a real definition.\footnote{We can learn this lesson from Plato’s Phaedo, in which Socrates concludes that concepts of universals, like “equality”, must have been learned in a previous life in which we were in direct contact with their Forms, given that we cannot now articulate their meanings. We do not need such a strong claim, but the point remains there is no good reason to assume that a reductive analysis of “probability” must be within reach if the term is to be used meaningfully.}

Moreover, none of the major reductive analyses of probability on offer are successful. “Epistemic probability” does not mean simply “degree of belief” (or “strength of belief” as in Hume), because our degree of belief can be mistaken. Therefore, not only is “degree of belief” not a reductive definition of epistemic probability, it isn’t even an accurate description. We can misjudge the epistemic probability of a possibility based on faulty reasoning. This is reflected in Aristotle’s distinction between “arguments that ought to persuade... and sophisms.” (Franklin 2002, 109) The fact that we judge something to be probable relative to our evidence does not mean that we judge correctly.

D.C. Williams writes:

Probability was thus for Hume... and later for De Morgan, only degree of belief... Nobody actually believes the subjective theory in any other sense than that he will state it when under the prod of epistemological debate. Its authors never actually hold, as their theory requires, that a person can improve his chances at a card game by drinking himself into an optimistic frame of mind. On the contrary, they use the objective calculus of probabilities as scrupulously as anybody. (Williams, 56)

“Subjective” is sometimes used to mean “epistemic”, sometimes “merely descriptive of psychology”. Epistemic probability is objective in the sense that it is fixed by objective epistemic norms, although it does not take account of physical facts beyond
one’s phenomenal awareness. Williams’ criticism is aimed at the “subjective theory” in the psychological sense.

The so-called “subjective” accounts of Bayes and Ramsey may well prevent optimism from hijacking one’s belief system in a game of cards, due to their requirements of consistency. But Williams’ target is not a Straw Man; Hume’s radical skeptic seems to hold that probability means mere degree of belief, with no such consistency constraint. In contrast, Williams points out that actual probabilities often come apart from people’s beliefs about them. A statement about the epistemic probability of something expresses but does not describe the speaker’s doxastic state (in terms of degrees). Just as “the door is open” does not mean “I believe the door is open,” “the door is probably open” does not mean “I believe the door is probably open.” It is a statement about the door, not about the speaker’s beliefs.

Williams rightly argues that probabilistic facts ground objective warrant for belief. But it seems that this is not a “real” (essential) or reductive definition of probability, either. It is because something is probable that the belief that it is probable is warranted, not vice versa. Williams writes, “In spite of a considerable tradition, logic itself is not a ‘normative’ science but a purely descriptive one…” (Williams, 30) By “descriptive”, he does not, of course, mean “psychological”, but “truth-preserving”. He explains:

The obligation to believe logical principles and to infer in accordance with logical principles is derived indirectly, from the fact that logical statements are true and that propositions logically inferable from true propositions must be true, and that it is good

---

47 This phrasing is due to Mark Crimmins.
that we should believe what is true. (Williams, 30)\textsuperscript{48}

Similarly, probable beliefs are warranted because it is good that we should take to be probable what is probable. So probability cannot be derived from or reduced to warrant for belief, as it seems that probability is at least as primitive.

Nor is probability relative frequency. The divergence is most apparent for small numbers of trials. For example, success on 1 out of 1 trial should not lead us to infer a probability of 1 for the next trial. Bayes, Laplace, and others provide precise formulas to calculate probabilities from frequencies given a number of trials, but each of these formulas contains an element of arbitrariness. For example, Bayes’ and Laplace’s formulas differ, and there may be no principled way to choose between them.\textsuperscript{49} Fortunately, exact numerical probability values are not needed to solve the problem of induction.

Like “objective warrant for belief,” “limiting relative frequency” is a correct description of probability, as Reichenbach held. But it is also not a reductive definition. Limiting relative frequency must be determined by counterfactuals, given that the world is finite. Such counterfactuals can only be made true by probability. Thus probability determines limiting relative frequency, not vice versa.

It seems likely that probability is conceptually primitive, with no possible reductive definition. Many philosophers have held this view. For example, Keynes writes, “Probability begins and ends with probability.” (Keynes, 322) Relatedly, Williams writes that probability relations are “pertinent to the truth values of their antecedents and consequents in the same way as are entailment and inconsistency.”

\textsuperscript{48} Thus, Williams believes that the normativity of logic is derived indirectly from “logical facts”.
\textsuperscript{49} See Dale 1999, pg. 61.
(Williams, 58) According to Williams, “Entailment and preclusion” are “the two extreme credibilities” along a continuum of credibility relations.” (Williams, 33-34) We can think of preclusion, or proof by contradiction, as a special case of inverse probability reasoning, in which probability 0 in the consequent entails probability 0 in the antecedent. \( P(e/H) = 0 \rightarrow P(H/e) = 0. \)

Keynes writes, “Of probability we can say no more than that it is a lower degree of rational belief than certainty.” (Keynes, 15) Similarly, Hume calls beliefs about probability “imperfect”: “Thus upon the whole, contrary experiments produce an imperfect belief.” (Hume, T1.3.12, pg. 185) Plato says that probability is “like truth.” (Franklin, 2002, p. 104) We may think of probability as a weakening of (warranted) certainty, with certainty as the primary notion. “Or we may make probability the more fundamental of the two and regard certainty as a special case of probability, as being, in fact, the maximum probability.” (Keynes, 15) This seems more natural, certainty being just one point along a spectrum of probability judgments. But regardless of which, if either, is more fundamental, “deductive logic and the theory of logical probability are in the same epistemic boat... an intuitive one.” (Stove, 176 and 185) Whereas Strawson holds that induction, like deduction, is known to be justified without argument, I claim here that the concept of probability used in induction is as free from the demand for conceptual analysis as the concepts used in deduction.

When events that are judged probable do not occur, this does not refute the judgment that they were probable relative to the information given at the time. “Sometimes the impression it makes is actually false... But the rare occurrence of this
kind... should not make us distrust the kind which ‘as a general rule’ reports truly.” (Sextus Empiricus 1957, I.175, p. 95) “A thing may very well happen in spite of the fact that some data render it improbable.” (Russell 1997) If the Sun suddenly and mysteriously disappears in ten minutes, it will still have been the case that, relative to our epistemic landscape as it is now, this will have been unlikely. “The proposition that a course of action guided by the most probable considerations will generally lead to success, is not certainly true and has nothing to recommend it but its probability.” (Keynes, 322) Due to the law of large numbers, the probability is very close to 1 that probable considerations will lead to success most of the time over large samples. However, over finite samples (even large samples), probability 1 is never reached through repetition. It is possible that cogent inductive arguments will lead, even in the finite long-run, to overall predictive failure. Probabilities cannot, therefore, be reduced to finite sets of actual or counterfactual happenings. The choice of an action on the basis of its probable outcome cannot be more simply explained: believing that something is probable does not involve believing that anything is actual; a consequence’s probability is (counterintuitively, perhaps) the final aim of our probability-based action.

The possibility that ‘probability’ is primitive is suggested by its cultural universality, as well as the difficulty of defining “probability” in simpler terms. Most often, perhaps, probability is understood in epistemic terms. The ancient Greek and Roman philosophers discussed probability, often in decidedly epistemic terms. Sextus Empiricus writes that the Academic called ‘probable’, “the apparently true presentation.” (Sextus Empiricus Against the Logicians, 1957, I.175, p. 95) Cicero
writes that sensations are probable when, “though not amounting to a full perception they are yet possessed of a certain distinctness and clearness, and so can serve to direct the conduct of the wise man.” (Cicero, 15) The appeal to guidance of judgment and direction of conduct might, or might not, suggest a leaning toward some kind of pragmatism. Aristotle gives a rough definition of probability in the Prior Analytics: “a probability is a generally approved proposition: what men know to happen or not to happen, to be or not to be, for the most part thus and thus.” (Aristotle 1989, 105). Probability is understood here as a matter of knowledge of frequencies. Though Aristotle first calls it a generally approved proposition, “know” implies that it is not just approved but epistemically worthy of approval. As Franklin says, probable reasoning, according Aristotle, “ought to persuade”.

Whereas Aristotle displays a focus on epistemic probability with the word “knowledge”, Cicero with “perception” and “sensation”, and Sextus Empiricus with “apparently” and “presentation”, Sun Tzu does not so clearly reveal whether he is working with an epistemic or objective notion of probability. Nevertheless, his use of the ‘probability’ concept is equally familiar.

2.1.3 Epistemic and Objective Probability

Our intuitive concept of probability can be used to refer to epistemic or objective probability, or both, often without explicit distinction. David Lewis writes that neither epistemic nor objective probability “can replace the other.” (Lewis 1980, 263) It seems that epistemic and objective probability are interdefinable, or at least inter-identifiable. Objective probability corresponds to epistemic probability relative to perfect knowledge, if perfect knowledge is possible. Wherever there is objective
probability, there is the hypothetical epistemic probability which aims to match it. Epistemic probability can diverge from objective probability only if and to the extent that there is ignorance.\textsuperscript{50}

Objective probability in the purest sense is exemplified in interpretations of quantum mechanics that deny physical determinism by rejecting hidden variables. Here probabilities are viewed as objective in that they are unlimited by lack of knowledge. These probabilities exist relative to all of the facts, not relative to a limited epistemic vantage point. Even if we knew everything there is now to know, we could not predict the future with certainty. Hume, Laplace and Einstein embraced determinism as an \textit{a priori} assumption, against pure, objective chance. For example, Hume writes that there is “no such thing as Chance in the world,” saying that we reason as if there were due to “our ignorance of the real cause.” (Hume, E6.46, p. 56)

Today, the question of whether the world is deterministic seems to belong less to philosophy than to physics, in virtue of such issues as the Einstein-Podolsky-Rosen Paradox, Bell’s theorem, and Bohmian mechanics. But we also talk about objective probabilities relative to different levels of description, and different ways of carving up our information. For example, biologists talk about the probabilities of observable traits relative to genetic information, as a function of observed frequencies. This probability is not relative to complete physical data, which together with Newtonian or Einsteinian equations almost completely determines the near future of all mid-sized objects, but relative to data that is relevant to knowledge that is practically attainable

\textsuperscript{50} Also, arguably, epistemic probability, relative to a state of incomplete knowledge, corresponds to objective probability relative to a hypothetical, partially complete, non-deterministic world constructed by mapping the content of the incomplete knowledge onto facts. But this is bound to be controversial and does not need to be accepted for our purposes.
by biological methods. Yet, given our limited knowledge of the world, we appropriately reason probabilistically nonetheless, relative to what we know or believe, or relative to fields of scientific inquiry that we believe we have good reason to believe we can trust.

A person acts on estimated probabilities relative to what she knows or believes. Russell writes, “probability is always relative to certain data.” (Russell 1997) Going a step further, Keynes writes, “When once the facts are given which determine our knowledge, what is probable or improbable in these circumstances has been fixed objectively, and is independent of our opinion.” (Keynes, 4) In this sense, epistemic probabilities are objective, rather than subjective in at least one of the traditional senses. They are objective facts about relations. This is true despite the convention of contrasting epistemic probability with “objective probability,” which is objective not only in the sense that, like the epistemic probability of an event, we can be mistaken about it, but also in the sense that it is relative to facts about the physical world, rather than relative to psychological states, as is epistemic probability.

An epistemic probability can arise from observations of regularity, which, in turn, are explained by an objective probability in the sense of a physical propensity. In such a case, if the person does not have an independent way of knowing about the objective probability’s continued existence, the epistemic probability will remain for some time even after the objective probability disappears. It remains rational for the person to believe that the regularity will continue given his/her ignorance of the objective reality. However, it is more epistemically likely, at any given time, given the belief that there is a physical propensity that explains/grounds the past regularity, that
this propensity will continue, because otherwise its own past persistence would be a colossal coincidence. (See Chapter 4.)

The only notion of probability directly relevant to Hume’s problem of induction is epistemic. “Hume does not include in our natural belief in necessary connection, especially in the scientific belief in the laws of nature, a metaphysical postulation going beyond our (Newtonian) epistemological practices.” (De Pierris 2002, 544) “The ‘sun’ to which he refers is anything you please - a mere dummy decorated with sunbeams.” (Black, 188) The problem of whether physical propensities can be inferred from our experience is different from the problem of induction; it is the problem of abduction. Unlike induction, in which we observe a pattern of ABABABABAB and after experiencing A we infer B in the next case\(^5\), we never observe the corresponding “Bs” in abduction. We infer that the world exists as more than our mere perception of it, but we never see the world as it exists independently of our perception – not even once, let alone repeatedly.

There is no metaphysical postulation, but our reasoning is the same as if there were. Hume suggests that the distinction between epistemic and objective probability judgments is irrelevant to the problem of induction, writing that it doesn’t matter whether we “suppose, that some secret causes, in the particular structure of parts, have prevented the operation. Our reasonings... and conclusions concerning the event are the same as if this principle had no place.” (Hume, E6.47, p. 58) Hume writes:

As to those impressions, which arise from the senses, their ultimate cause is, in my opinion, perfectly inexplicable by human reason, and 'twill always be impossible to decide with certainty, whether they arise immediately from the object, or are produc'd by the creative power of the mind, or are deriv'd from the author of our being. Nor is

\(^5\) The As and Bs are not necessarily perceptions of an external world.
such a question any way material to our present purpose. We may draw inferences from the coherence of our perceptions, whether they be true or false; whether they represent nature justly, or be mere illusions of the senses. Hume, T1.3.5.2, pg. 84

Even the Berkeleyan idealist runs up against the problem of induction. Establishing the epistemic probabilities that commonsense deems appropriate, not the physical or metaphysical bases of these probabilities, is the goal of a solution to Hume’s skeptical problem of induction. Epistemic probability, not objective probability, forms the justification of rational inductive prediction. Knowledge of the specific physical or metaphysical basis of the frequencies observed, if any, is not needed to infer probabilities with respect to unobserved cases.

Mackie holds that the same reasoning that justifies induction also justifies abduction, though he admits that the questions can be addressed separately:

I believe that ultimately similar answers can be given to both halves of the inductive problem. Nevertheless, the two halves can be distinguished: it would be one thing to justify the generalizing, the especially temporal extrapolation, another to justify the preference for this or that deeper explanation. (Mackie 1985, 161)

Even if you are agnostic about the nature or even the existence of a deeper explanation, epistemic probabilities are sufficient. It would be a mistake to rest a solution to a problem of induction upon a particular theory of what the deeper underlying explanation is. This would place an unnecessary burden on the solution. Armstrong and Foster, for example, do not sufficiently emphasize that their solutions to the problem of induction could withstand the destruction of their specific accounts of the nature of laws. (See Ch. 3)\textsuperscript{52}

\textsuperscript{52} However, Foster is aware of this, writing, “To say that a regularity calls for explanation is to say that it is rational to believe that it has an explanation independently of our knowledge of what the explanation is or is likely to be.” (Foster, 59.)
Mackie’s view involves a strong ontological commitment, although it is vague about what exactly it is committed to. Mackie writes:

On the hypothesis that all that is really there is four-dimensional machinery, the observed distribution of four-dimensional worms is a surprising coincidence; but on the rival hypothesis, which is hard to formulate but the point of which is that it finds some metaphysical truth in the thought of things as persisting through time and processes as projecting themselves in time, the general pattern of what we have observed is expectable... The inference to the suggested metaphysical truth as a deeper explanation is, then, a good inverse probability argument, and this metaphysical truth, whatever its precise formulation should be, would make it more reasonable to accept the principle of uniformity. (Mackie 1985, 175-176)

If there is nothing governing the behavior of matter in space-time, including its continuing existence (as four-dimensional worms) and motion, the various patterns that we observe across space-time are purely coincidental. Mackie appeals to the identity of objects over time as the metaphysical basis of the inferred “governing” probabilities, but is non-committal about what exactly this amounts to. But I think we should be equally non-committal, when proposing a solution to the problem of induction, about whether there is any such further metaphysical apparatus behind the epistemic probabilities. Perhaps there is nothing further to explain than that these epistemic probabilities hold. An inductive inference to support an epistemic probability is sufficient for solving the problem of induction, whether or not there is a deeper metaphysical or physical basis. I am not denying that there are powers or propensities in the world responsible for the regularities we observe, but if there are then the epistemic probabilities mirror them, and if not, the epistemic probabilities still obtain.
2.1.4 Criteria of Probability

Hume recognized three factors relevant to judgments of probability: indifference, frequency, and resemblance. Hume describes probability in terms of the principle of indifference, describing a 6-sided die, on which any four sides contain “a superior number of equal chances” to the other two. He also describes probability in terms of frequency, writing, “The first instance has little or no force: The second makes some addition to it: The third becomes still more sensible; and ‘tis by these slow steps, that our judgment arrives at a full assurance.” (Hume, T1.3.12) Thus, for Hume, our probability notion arises from both symmetry and past experience of regularity/patternedness. Additionally, he recognizes a role for degree of resemblance: “An experiment loses of its force, when transferr’d to instances, which are not exactly resembling; though, ‘tis evident it may still retain as much as may be the foundation of probability, as long as there is any resemblance remaining.” (Hume, T1.3.12)

These seem to be three distinct criteria for determining probability. But Hume collapses them. He asked, “...what is here meant by likelihood and probability?”, answering simply, “a superior number of equal chances.” (Hume, T1.3.11) Hume analyzes frequency in terms of indifference: “When in considering past experiments we find them of a contrary nature, this determination... offers us a number of disagreeing images in a certain order and proportion. The first impulse, therefore, is here broken into pieces, and diffuses itself over all those images.” (Hume, T1.3.12.) “Every past experiment may be consider’d as a kind of chance.” (Hume, T1.3.12) We are to imagine past instances like sides of a die, the throws of which will determine what happens next; higher frequencies correspond to a higher proportion of sides of
the die. Presumably, for the analogy to be complete, there would need to be an unknown number of unmarked “mystery” sides as well to account for possible outliers or unpredictable “black swans”.53 We can talk about the frequency of a characteristic within a population, or the frequency of a constant conjunction within our observations. Here Hume explains how the mind assimilates these two kinds of “frequency”. It is more difficult, perhaps, to see how probability judgments based on imperfect resemblance could be analyzed in terms of equal chances.54 Thus we seem to be left with multiple criteria for judging probability, suggesting that no single criteria for judging probability constitutes the essence of probability, while the intuitive simplicity of the concept of probability suggests that its essence is not constituted by the various criteria in tandem.

David Lewis claims that probability judgments should:

(1) obey the laws of mathematical probability theory; (2) avoid dogmatism, at least by never assigning zero credence to possible propositions and perhaps also by never assigning infinitesimal credence to certain kinds of possible propositions; (3) make it possible to learn from experience by having a built-in bias in favor of worlds where the future in some sense resembles the past; and perhaps (4) obey certain carefully restricted principles of indifference, thereby respecting certain symmetries. (Lewis, 1980, 289-90)

As I argue in Chapters 3 and 4, (3) is unnecessary. By avoiding dogmatism and reasoning consistently, it is possible to learn from experience without assuming prior to any observation of past regularity that the future resembles the past.

Sandy Zabell discusses Keynes’ concession to Ramsey that “the calculus of probabilities simply amounts to a set of rules for ensuring that the system of degrees of

53 See Nassim Taleb 2007.
54 For Hume, however, they are united by their common influence on the mind in terms of causing degrees of belief.
belief which we hold shall be a consistent system.” (Zabell, 131) But consistency is not the only norm on probability judgments; at the very least, we also need Lewis’s (2), avoidance of dogmatism. (See also the discussion of Mackie’s tolerance principle in Ch. 3.) Moreover, we can employ the principle of the supervenience of epistemic probability on psychological structure, which may or may not be considered a matter of consistency. (See Ch. 4)

2.1.5 Quantitative Probability, and the Principle of Indifference

The idea of probability existed long before its numerical expression in terms of fractions. Hacking writes that the Port Royal Logic of 1662 is “the first occasion on which ‘probability’ is actually used in what is identifiably our modern sense, susceptible of numerical measurement.” (Hacking, 2006, 25) It is amazing that even the simplest mathematical analyses of probability occur so late in history, despite the great usefulness of such analyses. But we should not be so blinded by this surprising fact that we exaggerate the novelty of our modern probability concepts and ignore the widespread, ancient use and cognitive importance of a non-quantitative notion of probability. Franklin criticizes Hacking’s assertion that “until the end of the Renaissance, one of our concepts of evidence was lacking: that by which the state of something can indicate, contingently, the state of something else.” He writes, “To reach these conclusions, it was necessary to ignore everything written about evidence in law and almost all medieval and early modern writing in Latin.” (Franklin 2001, p. 373.)

Laplace writes, “The theory of probabilities is at bottom nothing but common sense reduced to calculus.” (Laplace, 1986/1825, pp. 206-207.) Venn says that
probability is “a branch of the general science of evidence which happens to make much use of mathematics,” comparing the view that probability is “a portion of mathematics” to the view that geology is a portion of mathematics. (Venn 1866, Preface) In both cases, he writes, mathematics is “extraneous and accidental.” Although there is something to this, the comparison of probability to geology seems an exaggeration, and Venn admits that probability is more closely related to mathematics than most sciences.

Franklin holds a similar view, asking, “Whose business is it to explain the theory of probability?” He responds that “the modern answer” of mathematics “biases the theory towards aspects of probability that can be given precise numbers.” (Franklin 2002, 102) It is a mistake to assume that the concept of probability first arose when mathematics was applied to it. Numerical manipulation of probability is not an indispensable aspect of our probability concept, and much of our probabilistic reasoning is not subject to it.

Keynes argues that “the cases in which exact numerical measurement is possible are a very limited class, generally dependent on evidence which justifies a judgment of equi-probability by an application of the Principle of Indifference.” (Keynes, 160) Bayes/Price, Laplace, and many others take for granted the principle of indifference, or a uniform prior probability distribution, in their probabilistic reasoning from observed to unobserved. (See Ch. 3) It seems that Hume holds that all probability values are reducible to the probability 1/2. Hume writes, “It has been observ’d, that all single chances are entirely equal, and that the only circumstance, which can give any event, that is contingent, a superiority over another, is a superior number of chances.”
(Hume, T1.3.12) He writes, “The component parts of this possibility and probability are of the same nature, and differ in number only, but not in kind.” (T1.3.12) This position seems difficult to reconcile with his view, discussed above, that “an experiment loses of its force, when transferr’d to instances, which are not exactly resembling.” It is not at all obvious that degrees of resemblance can be broken down into exact quantities by way of the principle of indifference.

In everyday life, we are usually unable to assign exact numbers to the probabilities of events we deem likely or unlikely. Keynes writes of civil law cases in which a distinction is made “between probabilities, which can be estimated within somewhat narrow limits, and those which cannot.” (Keynes, 24) In such cases, “relative to the completest information available... the probability could be by no means estimated with numerical precision.” (Keynes, 27) Similarly, people believe that the Sun will probably rise tomorrow, but this does not commit them to any exact numerical assignment. Even in science, “probability is still far from fully quantified.” (Franklin, 2002, 362) Strawson goes further: “In fact, we can never describe the strength of evidence more exactly than by the use of such words as ‘slender’, ‘good’, ‘conclusive’, &c.” (Strawson 1952, 247) Ironically, he uses this to argue against the sampling solution to the problem of induction: “Use is to be made of the arithmetical calculation of chances,” but “the prospect of analysing the notion of support in these terms seems poor.” (Strawson 1963, 252) Here Strawson overlooks the possibility of non-quantitative forms of the sampling solution.

It might be argued that the vagueness of the probabilities considered in law, science and everyday life results merely from limited skill in probabilistic reasoning.
However, even in principle, it may not always be possible to assign exact numerical values to probabilities, no “theoretical rule” having “ever been suggested”. (Keynes, 27-28) The criteria are too complex for an exact analysis: “We cannot always weigh the analogy against the induction, or the scope of the generalisation against the bulk of the evidence in support of it.” (Keynes, 30) Pace Laplace and Bayes, who illustrated their theories with idealized examples in which the principle of indifference can be cleanly applied, Keynes writes that in the real world, “A degree of probability is not composed of some homogenous material, and is not apparently divisible into parts of like character with one another.” (Keynes, 30) Certainly, at least, we are rarely aware of any complex applications of the principle of indifference that might underlie our probability assessments in everyday life. This does not mean that numbers cannot be applied at all to most real life situations in which probability is utilized. Keynes writes, “Many probabilities, which are incapable of numerical measurement, can be placed nevertheless between numerical limits.” (Keynes, 160) We can often assign a range, but rarely an exact probability. Vagueness is unavoidable in many probabilistic situations.

Moreover, we can misjudge even the vague boundaries of our epistemic probabilities. “Unknown epistemic probability” may seem paradoxical, for how could we fail to know the probability of something relative to what we know? Nevertheless, this is both possible and commonplace. When first asked the probability that $x$, we often make an estimate, which upon further reflection we realize should be revised, due to more careful consideration of the same information. Keynes asks whether “by saying that a probability is unknown,” we mean “unknown through lack of skill in
arguing from given evidence, or unknown through lack of evidence?” He answers that “The first is alone admissible.” (Keynes, 31) If it were through lack of evidence, we would not be dealing with epistemic probability, i.e. probability relative to what we know, for our knowledge does not contain the missing evidence. Rather, “The evidence justifies a certain degree of knowledge, but the weakness of our reasoning power prevents our knowing what this degree is.” (Keynes, 32)

But this leaves us with a difficult question. As we have said, the epistemic probability of any event is unknown to the extent that we lack skill in reasoning relative to what we know. This lack of skill is relative to some paradigm of logical reasoning. But it is not relative to perfect logical reasoning. “The degree of probability, which it is rational for us to entertain, does not presume perfect logical insight.” (Keynes, 32) If it did, the epistemic probability of any provable mathematical theorem would be 1, even if it were unproven. But this is not what we mean, intuitively, by epistemic probability. The epistemic probability, relative to us today, of the truth of an unproven theorem, such as the Riemann hypothesis, is not 0 or 1 depending on whether it is true, but perhaps roughly 85%. Thus the hypothetical level of skill in reasoning relative to which epistemic probability is determined, the golden mean between complete inferential incapacity and perfect recognition of all of the obscure implications of one’s beliefs, remains unspecified.55

An important question concerns the “default”: whether differences are presumed irrelevant until proven relevant, or vice versa. How does one know when one is presented with two equal, single chances? For our purposes, we can afford to be

55 Perhaps we should say that there are many different epistemic probabilities, relative to different ideals of rational credence. (This was suggested by Mark Crimmins.)
conservative and only endorse the principle of indifference in cases in which there are no possible grounds for distinguishing the probabilities of two possibilities, as in cases of exact resemblance, including Groundhog Day.

Solutions to the problem of induction commonly invoke the principle of indifference in two capacities: first, to justify the prior assumption that, at any given time, the universe is as likely to be any one way as any other (where the space of possibilities is divided, question-beggingly perhaps, in some “natural” way); second, to justify the prior assumption that all possible degrees of probabilistic dependence among the probabilities of events at different times are deemed equally likely. We can understand this by analogy with a die which may or may not be weighted. Before rolling the die for the first time, we use the principle of indifference, first, to reason that it is equally likely to land on each side, given our ignorance of the weight distribution; second, to reason that any given weight distribution is equally likely.\(^{56}\)

Both of these uses of the principle of indifference can be replaced by a weaker principle of tolerance, like the one adopted by Mackie. (See Ch. 3) We do not need to assign equal probability to all possible states and degrees of dependence. We only need to assign them non-zero probability. As long as it is not effectively ruled out \textit{a priori} that an event will occur (such as the Sun rising tomorrow) or that probabilities over time are interdependent, observations of regularity in the past will increase the probability that probabilities over time are dependent, and that the regularity will continue for some length of time.

Someone might believe that the \textit{a priori} probability that the Sun will rise in the

\(^{56}\) This discussion of the principle of indifference is closely related to the discussion of the biased sampling objection in Ch. 4.
future, or that observations at different times are probabilistically dependent, is extremely low. Although I do not see the motivation for this view (it seems *ad hoc*), it does not need to be rejected by the solver of the problem of induction. Increased probability can accrue as long as a non-zero probability is assigned to dependence and the Sun rising in the future.

Additionally, although nothing as strong as the principle of indifference is needed to solve the problem of induction, some applications of the principle of indifference nevertheless seem justified. A common objection to the principle of indifference is that different, seemingly equally principled ways of dividing up a sample space yield inconsistent probability judgments, leading to contradiction. For example, we might say that a basket holds red, orange, and green balloons, or instead say that it holds reddish (red or orange) and green balloons. (Perhaps, we might surmise, the orange balloons are red ones whose colors have faded.) If we describe it in the first way, the principle of indifference implies that the probability of choosing a green balloon is 1/3; if in the second way, it implies the probability is 1/2. Similarly, “Should one assign a probability of 1/4 to each of a parabola, circle, ellipse, or hyperbola when considering the orbit of a comet, or should a circle and a parabola receive smaller probabilities since they form special cases of ellipses and hyperbolas?” (Eberhardt and Glymour, in Reichenbach 2007, p. 5) Since there is no way to decide, it seems that neither answer is uniquely correct.

But this criticism does not apply to every version of the principle of indifference. The problem only arises if we are liberal in applying the principle of indifference. Hume, for instance, seems to endorse a very liberal form of the principle
of indifference: “Where nothing limits the chances, every notion, that the most extravagant fancy can form, is upon a footing of equality.” (T1.3.11) But lacking knowledge of relevant differences between two scenarios with respect to probability may not be enough; we may instead need to see clearly that there is a lack of any relevant difference, in order to be able to justifiably appeal to the principle of indifference. Thus Mackie writes that the criticism that there are different ways of dividing sample spaces only applies “where a range of possibilities does not divide unambiguously into similar alternatives.” (Mackie 1979, 168) Keynes says the same: “The paradoxes and contradictions arose, in each case, when the alternatives, which the Principle of Indifference treated as equivalent, actually contained or might contain a different or an indefinite number of more elementary units.” (Keynes, 59) For example, modifying Oresme’s example that the number of stars is probably not a cube, it seems a justifiable use of the principle of indifference, not relying on induction, to conclude that the probability that the number of stars is even is 1/2. 

Here is another, related example. Imagine that you are given the information that exactly one of two sentences you are about to be read is true. The sentences are given arbitrary names, e.g. “A” and “B”. It seems that we are justified in using the principle of indifference to argue that the “prior” probability that each sentence is true is 1/2. The example resembles but is importantly different from one rejected by Keynes:

57 Moreover, Mackie shows that the principle of indifference can be relaxed, appealing instead to a “principle of tolerance,” a non-zero prior probability that probabilistic dependences hold over time: Whether the balloons are reddish and green or red, orange and green, there is a non-zero probability that any particular balloon is green.
Jevons’s particular example, however, [“A Platythliptic Coefficient is positive”] is also open to the objection that we do not even know the meaning of the subject of the proposition. Would he maintain that there is any sense in saying that for those who know no Arabic the probability of every statement expressed in Arabic is \( \frac{1}{2} \)? How far has he been influenced in the choice of his example by known characteristics of the predicate ‘positive’? (Keynes, 43)

Jevons’ and Keynes’ examples are similar to my example of the two arbitrarily named sentences one of which is known true, but the latter is a better candidate for a situation in which the principle of indifference can be cleanly applied. That is, it is easier to argue that the probability that a sentence is true is \( \frac{1}{2} \) if we know, as in my example, that it is one of two arbitrarily ordered sentences exactly one of which is true. For it may be, for example, that we have some reason to believe that most sentences are false, most possibilities being unrealized, or that most things are not positive (whatever a Platythliptic Coefficient might be).

Now, imagine that you read one of the two sentences (of which you know in advance that exactly one is true), and you are almost certain that it is false. Although you have not read the second sentence, you have grounds for believing that it is probably true. This is true though you do not know the content of the sentence that you judge probable. This is “near-proof by near-contradiction”: an argument that the probability of a hypothesis is increased in virtue of being the only alternative to (i.e. the logical complement of) an extremely unlikely hypothesis.

Next, imagine that before reading either sentence, you think it is much more likely that sentence A is false, perhaps because you have heard the opinion of someone who has read sentence A. But you then read sentence B, and believe independently that it is probably false. Even though your prior probability assignment of A is low, it is raised by learning that the probability of B is low. We only need to avoid arbitrarily
assigning a zero prior probability to either of the two unknown sentences before we read them, in order for a learned low probability of one to imply an increased probability of the other. As long as we are not initially certain that sentence A is false, learning that B is probability false increases the probability that A is true.

This example illustrates that such arguments can be persuasive even in cases in which the specific content of the alternative hypothesis is unknown (it may only be referred to as the alternative), in cases in which its prior probability is vague, indeterminate, or unknown, and in cases in which the prior probability is low, as long as it does not have prior probability 0. We do not need the principle of indifference; mere tolerance in Mackie’s sense is sufficient to solve the problem of induction. Moreover, we can have evidence for a hypothesis, which we refer to by an accidental property (such as its status as an “explanation” of past regularity in nature or as an “alternative” to the Regularity theory of causation), without knowing or taking a position on what it is specifically that we are hypothesizing or how likely we should otherwise have thought it to be.

2.1.6 Probability and Cross-Induction

Defeating conditions change the appropriate reference class for induction. Keynes writes, “If the acquisition of new knowledge affords us additional relevant information about the particular instance, so that it ceases to be a random member of the series, then the statistical induction ceases to be applicable.” (Keynes, 411-412) If we learn that a building is scheduled to be torn down tomorrow, our knowledge that it has been standing for many years does not lead us to deny that it probably will be torn down tomorrow. The population over which frequencies are to be noted changes. We change
the reference class from “longstanding object” to “longstanding object scheduled to be torn down the next day.” This is a case of “cross-induction”, a clash of inductive arguments relative to different reference classes. In this case, though not in all, one reference class trumps the other.\footnote{Resilient theories rely on reference classes that are not easily trumped.} In other cases, frequencies relative to different reference classes will jointly contribute to the overall inductive probability.

Where the future will probably not resemble the past in some respect, it is only because ending one pattern is necessary for continuing a more extensive or resilient pattern. For example, the world polar bear population has been steadily declining, now at its lowest level in 200,000 years. We can use this information to inductively infer either that the population will continue to decrease in the near future, given that this is a trend, or that the population will not continue to decrease, as its being greater or equal to its current level is also the historical norm. Here it is clear which side of the cross-induction takes precedence. The population is more likely to continue declining for some time, because this state of affairs would continue more patterns overall.

\subsection*{2.2 Coincidence}

Foster defines the everyday notion of a coincidence as an occurrence that is unlikely and “independently marked out” or independently salient or significant. (Foster 2004, 67. Hereafter, let us substitute “marked” for “independently marked out.”) Similarly, Hart and Honore define a coincidence as “The conjunction of two or more events in certain spatial or temporal relations” which are “very unlikely by ordinary standards,” “for some reason significant or important,” “occur without human contrivance,” and “are independent of each other.” (Hart and Honore, 78) “Occur without human
contrivance” is redundant, since this is implied by their independence. Moreover, if they were not independent, they would not be unlikely, so this too is redundant. This definition therefore reduces to Foster’s, replacing “independently marked out” with “for some reason significant or important.”

Diaconis and Mosteller also give a definition which can be interpreted to mean the same. First they offer, as a working definition of coincidence, “a surprising concurrence of events, perceived as meaningfully related, with no apparent causal connection.” (Diaconis and Mosteller, 2) But they admit that “a more liberal definition is possible,” one not emphasizing “the observer’s psychology,” implied by the terms “surprising, perceived, meaningful, and apparent.” (Diaconis and Mosteller, 2) Reinterpreting “surprising” and “meaningful” in non-psychologistic terms, and recognizing that a causal connection would prevent a correlation from being unlikely, we are left with something very much like Foster’s criteria of being unlikely and independently marked out.

Moreover, we can understand the property of being marked as the property of being marked by a natural relation, in Hume’s sense. A natural relation exists between ideas iff an idea of one naturally introduces to the mind an idea of the other. In contrast, ideas that are related merely by a contrived “philosophical relation” do not naturally introduce each other to the mind, but only by an arbitrary act of putting together these ideas in the mind.

Correspondingly, let us distinguish a strong and a weak sense of ‘unlikely’.

---

59 Mark Crimmins asks whether we have not two distinct senses of ‘unlikely’, but different contexts in which the same concept is employed. This possibility is suggested by the fact that we do not seem to have two corresponding senses of ‘likely’. (See below.) Perhaps it is merely the epistemic state relative
We can illustrate this distinction with the example of the blade of grass. Mary picks a blade of grass from a lawn. Jim jokes that it was unlikely that she would have picked that very blade among the hundreds she was equally likely to have picked, pretending to be surprised. But there is nothing special about the blade of grass she picked, other than the fact that it is the blade she picked. It was unlikely that she would have picked the blade that she picked (*de re*), but it was certain that she would have picked the blade that she picked (*de dicto*). It is clearly not a coincidence that she picked the blade that she did. A coincidence is unlikely in the strong sense; Mary picking the blade she picked (*de re*) is unlikely in the weak sense, because it is not marked out independently of being picked.

‘Coincidence’ in common usage, and as used here, means unlikely in the strong sense. Something is unlikely in the strong sense if it is unlikely in the weak sense as well as marked. Virtually every actual occurrence is unlikely in the weak sense; the chance that each snowflake will fall exactly where it does is astronomically low. And so, contradictory though it may seem, unlikely occurrences (in the weak sense, unlike coincidences) are, in any given case, likely.

Jim does not refer to the blade of grass as the blade of grass picked by Mary, but rather, as “that blade of grass”. Jim’s joke plays on the ambiguity of the reference class; in particular, in the ambiguous mode of presentation of “that”. If the mode of

to which something is unlikely that changes, not the meaning of ‘unlikely’. Nevertheless, I will continue to talk loosely of two senses of ‘unlikely’ as a shorthand for the two contexts of use.

If a state is likely, it does not make a linguistic difference whether it is marked or unmarked; there is no special name for a likely event that is marked. If an event is likely and marked, it is caused (or law-governed, or some such). If it is likely and unmarked, it is also caused. If likely and marked, the cause reduces the coincidence of the markedness. If likely and unmarked, the cause does not reduce the coincidence of the unmarked state in itself, because the unmarked state would not be a coincidence even if uncaused. Instead the relation is causal in this case in virtue of avoiding another coincidence, in terms of the conjunction of the intrinsically unmarked state with the earlier state taken to be its cause, relative to which it is probable.
presentation of "that" includes being mentioned by Jim because it is the blade of grass in Mary's hand, then it is not at all unlikely that she would have picked "that" blade of grass. It is unlikely relative to one reference class, not relative to another, and in particular, not relative to the reference class that would justify surprise, that Mary picked "that" blade of grass.

If Jim describes the blade of grass *qua* blade of grass that he chooses to refer to, the probability of Mary picking "that" blade of grass is 1, given that Jim deliberately chose to remark about the blade of grass that Mary picked. The blade of grass is not marked out independently of being picked by Mary; Jim's remark is "biased" in that there is a relation of probabilistic dependence between his method of choosing a blade of grass to remark on and the blade of grass that Mary picked. It is true that it is unlikely, in a sense, that Mary would pick that blade of grass, if we ignore the fact that we are talking about it because Mary picked it (if we include this information, then it is no longer unlikely that it would be the blade that Mary picked). The chances are miniscule, from some perspectives. But other than being the blade picked by Mary, the blade is not marked, so it is unlikely only in the weak sense. It is referred to because she picked it, so it is unlikely that she would have picked it only in the weak sense, in which we ignore the known common causal origin of Jim's question about its likeliness with its actual occurrence. Therefore it is not surprising and not the kind of thing that ought to make us suspect a coincidence-lessening explanation operating beneath the surface. The events are not independent, and therefore the event does not constitute a coincidence. It is a causal regularity, not a coincidence. Had it been otherwise, had they pointed at the same blade of grass
independently, it would then be a coincidence: unlikely in the strong sense.

The definition of coincidence as unlikely in the strong sense is the same as Foster’s definition of a coincidence as unlikely and independently marked out (i.e. “interesting,” “meaningful,” significant, natural, simple, or salient), which corresponds to the common everyday meaning of the term. For example, it is a coincidence when friends meet in a foreign city by chance, because it is unlikely that this will happen in any particular instance, and it is independently marked out by the fact that ‘being friends’ or ‘knowing each other’ is an interesting, natural, simple property that stands out in an objective sense, much more than ‘being very distant cousins’. While most or perhaps all events fall under at least one description such that they are weakly unlikely, the unlikely property is not, in most cases, independently marked out, and therefore it is not unlikely in the strong sense, i.e. not a coincidence.

David Owens means by a ‘coincidence’ any uncaused conjunction (perhaps the same as what Hume calls a “philosophical relation” in T1.1.4). He writes, “The conjunction of my now driving a green car and the Queen’s beginning a visit to France in exactly a year’s time is a coincidence.” (Owens, 7) It is an Owens coincidence if Bill Clinton lives to the age of 84 years, the conjunction of a person’s age and an arbitrary number. There is no “anchor”, nothing special about this particular number, so it is not a coincidence (not unlikely in the strong sense). But it is unlikely, in the weak sense, that Bill Clinton would live to this exact age. On the other hand, it is very likely, even certain, that Bill Clinton would live to some age which it was unlikely that he would live to. This is why we say it is not unlikely in the strong sense, though it is unlikely in the weak sense. It would be a minor coincidence, on the other hand, if Bill
Clinton dies at the age of 100 years, not because there is anything *naturally* special about the number 100 but because specialness is bestowed on it by the common convention of using the base-10 number system. This is a coincidence because it is a conjunction of a famous person’s age at death and the conventions that make the number 100 special. It is unlikely because causally unrelated to the conventional specialness of the number, and marked due to resemblance (sameness of number).

The strong sense of unlikely includes John Foster’s idea of being “independently marked out,” a feature necessary to being a coincidence. It is unlikely in the strong sense if you get four aces in the first round of poker (but only in the weak sense if you get some uninteresting assortment, like 5, Jack, 2, 9, King), if you drop the deck and the cards form the shape of a walrus, or if all of the air particles in the room, by chance, assemble in one corner. For this reason, hypotheses that lessen coincidence overall are more likely, other things being equal. Our explanations are more likely to be correct if we aim to reduce coincidence as much as the observations allow.

Just as we can distinguish weak and strong senses of “unlikely,” the latter being a “coincidence, we can distinguish weak and strong senses of “coincidence,” the latter being a colossal coincidence. It is likely that Bill Clinton will live to some (weakly) unlikely age, such as 84, and unlikely that he will live to some (strongly) unlikely age, such as 100. Correspondingly, it is likely that (weak) coincidences will sometimes occur, but unlikely that (strong) coincidences will ever occur. “The law of truly large numbers” states, “With a large enough sample, any outrageous thing is likely to happen.” (Diaconis and Mosteller, 14) But with any finite sample, there will
always be something so outrageous that it is unlikely to happen.

A strong coincidence can be called a “colossal coincidence”; it is a coincidence larger than should be expected to ever occur. A coincidence is, by definition, unlikely to occur in any given case, but it is not unlikely to occur anywhere in a large universe/sample. In contrast, a colossal coincidence, or a coincidence in the strong sense, is unlikely to occur even in a large finite universe/sample. “Colossal” is defined relative to the size of the universe; the larger the universe/sample, the more unlikely something must be in order to qualify as a colossal coincidence. Given the size of the universe or sample, a colossal coincidence is a coincidence larger than we should expect to occur anywhere within it. In contrast, weak coincidences should be expected to occur sometimes, but not in any given case (chosen independently of markedness).

Coincidences, but not unlikelihoods in the weak sense, are unlikely to occur in any particular case. Because coincidences are unlikely to occur in any particular case, we have reason to be suspicious of an alleged coincidence, except where we have reason to believe that we are focusing on a case because of its perceived significance (a form of anthropic reasoning). Coincidences are common in everyday life. Yet, though common, they are relatively infrequent over the long run, and if they seem to be occurring too frequently, there is probably a coincidence-reducing explanation. Any particular case is unlikely to be a coincidence, ceteris paribus. It is due to this definitional unlikeliness that we know, in every day life, to be suspicious of coincidences.60

Such inference to non-coincidence reasoning is common in science,

---

60 Suspicousness of coincidence may be responsible for at least some forms of Ockham’s Razor. If we complicate a hypothesis needlessly, it is more coincidental if it turns out to be correct.
philosophy, and everyday life. For example, Aristotle writes that what is “everywhere, and in all things,” does not “proceed from fortune.” (Aristotle, On the Heavens 289b27-28, in Taylor) Cicero says that “this elaborate and beautiful world” could not be produced by the “fortuitous collision” of particles,” any more than the Annals of Ennius could be produced by letters “thrown together into some receptacle and then shaken out on to the ground.” (Cicero, 213) Salmon concludes a similar argument, “Even though a chance coincidence of this sort is possible, it is too improbable to be accepted as such.” (Salmon, 1998, p. 96) Sober writes that when a hypothesis of dependence “renders the observations more probable than does the hypothesis of Mere Coincidence... the evidence favors the first hypothesis over the second.” (Sober, p. 8.)

Dependence is more likely because without it, our observations would be less likely.

Similarly, it either is or isn’t a coincidence that the Sun has risen with such regularity. If it isn’t a coincidence, it is likely to rise again. It is unlikely to be a coincidence. Therefore, it is likely to rise again. More generally, it either is or isn’t a coincidence that property P is correlated with property Q in our experience. If it isn’t a coincidence, the properties are likely to be correlated in what we have not experienced. (See Ch. 4 for further explanation)

One might object that people commonly commit a fallacy of “over-surprise,” believing experiences to be more unlikely than they are, making us erroneously attribute common causes where no such explanation is needed. For example, Skinner (1941) criticizes the common belief that Shakespeare deliberately uses alliteration, arguing that the extent of alliteration matches expectation on chance. Similarly, Gould (1988) criticizes the common belief that professional baseball and basketball players
have “hot streaks.” It seems that this tendency to underestimate randomness is due at least partly to our tendency to inadequately integrate the statistical implications of the fact that we notice many different kinds of matches, and near matches - such as friends having the same first name, the same birthday, or birthdays within a few days of each other. (Diaconis and Mosteller, 10) However, it seems that this fallacy can be avoided with careful reflection, and does not disprove the cogency of inference to non-coincidence reasoning in general.

Although we usually think of coincidences in terms of resemblance, it seems that they can sometimes be manifested instead in terms of contrariety. Sober gives an example illustrating this: “If you and I always order different desserts when we dine together at a restaurant, the waiter may be right to suspect that this is not a coincidence.” (Sober, online) It seems that this coincidence is a correlation not of resemblance, but of difference; that is, the significance attending it is due to contrariety. However, even in this case, coincidence relies ultimately on resemblance, in terms of the consistency of us ordering different desserts. Perhaps all significance is due to resemblance, at some level of analysis.

2.3 Explanation

There are different kinds of explanation, corresponding to different kinds of understanding. Sometimes an explanation is called for in the sense of a mere description of some (any) causal chain that lead to the state of affairs to be explained. At other times, we seek, in the course of describing how something came to be, to

---

61 While these particular examples are controversial, the broader point that we have a tendency to see order in randomness - like constellations in the stars or faces in our Cheerios - is more generally acknowledged.
thereby also demonstrate that it is not as coincidental as it might seem, *prima facie*. We seek not just to be able to say something about the generation of the explanandum, but to explain how this observed explanandum is probabilistically plausible.

Inference to the best explanation is common in philosophy. For example, Descartes writes, “Since there is no other equally suitable way of explaining imagination that comes to mind, I can make a probable conjecture that the body exists.” (Descartes, 51) It is debatable whether all inference to the best explanation is inference to non-coincidence. If not, I am concerned here only with inference to the best explanation in the sense of inference to non-coincidence, i.e. a plausible-probability explanation. Regarding this type of inference, Salmon writes, “The principle of common cause states, roughly, that when apparent coincidences occur that are too improbable to be attributed to chance, they can be explained by reference to a common causal antecedent.” (Salmon 1988, 97) Where the only alternative is colossal coincidence, we should expect a common cause.

Relatedly, Hempel writes that given an explanation (*explanans*) of an occurrence (*explanandum*), we have reason to believe that the occurrence “was to be expected.” Knowing the explanans helps predict the explanandum, providing evidential support. But knowing the *content* of the explanans (as opposed to knowing merely that there exists some unknown explanans) is only sufficient, not necessary, for having grounds for prediction. Knowing that there exists an explanans suffices even if you do not know the specific content. “Whereas an understanding of a phenomenon often enables us to forecast it, the ability to forecast it does not constitute an understanding of a phenomenon.” (Scriven, 54) The length of the shadow helps
predict but does not explain the height of the flagpole, and the position of the barometer needle helps predict but does not explain changes in atmospheric pressure; the height of the flagpole and weather conditions are causally prior to the length of the shadow and position of the needle. (See Bromberger 1966, p. 92) However, the length of the shadow and the height of the flagpole (as well as the weather conditions and needle position) are held to be causally interdependent, which gives us reason to predict the shadow length based on the flagpole height. We do not need to know what the causal relationship/explanation is in order to know that there is an explanation that justifies prediction.

This is related to the point that “We can infer the truth of an explanation only if there are no alternatives that account in an equally satisfactory way for the phenomena.” (Cartwright, 76) People often try to give more detailed explanations than their evidence justifies. However, in such cases, even if the details of our explanation are incorrect or improbable, our hypothesis that there is some explanation is often correct.

Consider an example of a situation in which the details of an explanation were incorrect but the underlying assumption that there was some explanation to be found was correct, from a New Yorker article discussing a Powerball lottery drawing with more than 100 winners. “Lottery officials suspected a scam until they traced the sequence to a fortune printed with the [winning] digits “22-28-32-33-39-40” along with the prediction: “All the preparation you’ve done will finally be paying off.” (http://www.newyorker.com/archive/2005/06/06/050606ta_talk_olshan#ixzz0kRyZzJB) The particular explanatory speculation that it was a scam was incorrect, but the
more general inference that there was *some* coincidence-lessening explanation to be found was correct. This example demonstrates the wealth of possible common causes of a correlation, some of which may require imaginative skill to consider.

Here is another example of a correlation with a presumed but unknown explanation. Of the 60 quarters with state emblems that I have recently collected, 10 of them, all minted in the year 2002, represent North Dakota. I have no explanation, but I have strong reason to suspect that an explanation exists, i.e. to believe it is not purely a coincidence. Here the links between explanation (in the relevant sense), coincidence and probability are intuitively manifest.

Not all inference to the best explanation is inference to lesser coincidence. But much of it is. Being a coincidence, being improbable (in the strong sense), and being unexplainable (in the sense of a plausible-probability explanation), all mean the same. As David Owens writes, “The more of a coincidence an event is, the less amenable it is to explanation.”62 (Owens, 13) For example, if we explain the fact that my brother and I are at the bakery at the same time by appealing to the fact that we walked there together, this provides an explanation by eliminating the initial appearance of coincidence. In contrast, if there is no such explanation of the fact that we are there at the same time, because the causal chains responsible for each of us being there are not related in the right way, then it is a coincidence that we met there. Moreover, we are more likely to accidentally meet at a bakery in our home town than on independent trips to Greenland. The latter is more of a coincidence, because more unlikely, though equally “independently marked out” as the co-incidence of the location of brothers.

---

62 This applies to “coincidence” in both Owens’ sense and the sense employed here, which also includes being independently marked out.
Bas van Fraassen challenges the link between “coincidence” and “explanation”, writing, “It is illegitimate to equate being a lucky accident, or a coincidence, with having no explanation.” (Van Fraassen, 25.) He gives the example, “It was by coincidence that I met my friend in the market—but I can explain why I was there, and he can explain why he came, so together we can explain how this meeting happened.” Similarly, Hume argues in the voice of Cleanthes that an explanation of the parts is sufficient for an explanation of the whole. (Hume 1993, 93) David Owens objects that it is “wrong to think that explaining the existence of each particle suffices to explain the existence of them all.” (Owens, 14) He writes, “To explain each of the parts of an event is not necessarily to explain the whole event. To explain the whole we must show that its parts share a common cause.” (Owens, 13) This seems right, but further qualification is needed. Suppose that I met my friend in the market, and there is a common cause of us both being there; perhaps we saw the same TV commercial. Although this is an explanation in a sense, and even an explanation in terms of a common cause, it is still a coincidence that we were influenced by the same commercial. Therefore, for the whole to be explainable in terms of a coincidence-lessening explanation, it is not sufficient that the parts share a common cause, for the very fact that they share a common cause might be an equally great coincidence.

The difference of opinion over the meaning of “explanation,” between Owens and Salmon, on one hand, and Hume’s Cleanthes and van Fraassen, on the other, is a mere “semantic debate”, a dispute based on an ambiguity. There is a sense in which an explanation of the parts counts as an explanation of the whole, as Hume and van Fraassen hold, and another sense in which it does not. An explanation of the parts does
not necessarily reduce the coincidence of the whole. Cartwright writes, “the argument from coincidence supports a good many of the inferences we make to best explanations.” (Cartwright, 75) Often, when we ask for an explanation, in both everyday life and science, we seek an explanation in the sense that lessens coincidence. To give a mere description of causal origins when asked for an explanation in this sense would be to mistake the question.

Just as van Fraassen argues that being a coincidence does not entail having no explanation, Glymour and Railton write that being unlikely does not entail having no explanation. Glymour writes, “One can explain unlikely outcomes just as well as one can explain their more probable alternatives.” (Glymour, 138) Railton gives an example:

Why should it be explicable that a genuinely random wheel of fortune with 99 red stops and 1 black stop came to a halt on red, but inexplicable that it halted on black? Worse, on Hempel’s view, halting at any particular stop would be inexplicable, even though the wheel must halt at some particular stop in order to yield the explicable outcome red. (Railton, 123-124)

Like van Fraassen, Glymour and Railton trade on the ambiguity in “explanation” between a description of how something came to be and a description that lessens coincidence. It is unlikely, but not a coincidence, if the wheel halts on any particular one of the 99 red stops, but it is a coincidence if it halts on the 1 black stop because it is independently marked out as the only stop that isn’t red. The sense of explanation that van Fraassen, Glymour, and Railton invoke in these passages is distinct from the sense appealed to in inference to coincidence-lessening reasoning in science and everyday life.
In summary, coincidence can be identified with a sense of “unlikely,” which includes being independently marked out. Much of inference to the best explanation is inference to lesser coincidence, an *a priori* form of reasoning.
Chapter 3. Critical History of the Solution

3.1 Introduction

Keynes expresses the essence of the solution:

D’Alembert, Daniel Bernoulli, and others... showed... that certain observed series of events would have been very improbable, if we had supposed independence between some two factors ... and they inferred from this that there was in fact a measure of dependence or that the occurrence had probability in its favor. (Keynes, 370)

It is very unlikely that unlikely events will, by chance, frequently co-occur. Where events co-occur with a frequency much greater than statistical expectation based on an assumption of probabilistic independence, dependence is more likely. Does this argument imply that dependence is more likely in the past alone, as Hume held, or does it also imply that dependence is more likely in the unobserved near future? We shall return to this question later.

Mackie makes a similar argument: “because some result which the falsity of a certain hypothesis would render improbable has been observed, it is now likely that that hypothesis is true.” (Mackie 1979, 168) If something is very unlikely unless something else is true, and we learn that the former is true, it is more likely that the latter is true. For example, it is statistically extremely unlikely that someone would win the lottery 10 times, having purchased only 10 tickets in his lifetime, without some kind of foul play or error in the system causing probabilistic dependence among the different instances. Therefore, if someone wins the lottery on 10 out of 10 attempts, it is likely that some such coincidence-lessening, non-chance “explanation” is responsible. Similarly, if it weren’t the case that vast regularities in progress are

---

63 “I ask, why in other instances you presume that the same power still exists, merely upon the appearance of these qualities? (Hume, T1.3.6, 91)
likely to continue, they would be extremely unlikely to arise.\textsuperscript{64}

3.2 Bayes and Price

The same point can be made using Bayes’ Theorem:

\[
P(H|E) = \frac{P(E|H)P(H)}{P(E)}
\]

Dividing both sides by P(H), this implies:

\[
\frac{P(H|E)}{P(H)} = \frac{P(E|H)}{P(E)}
\]

This implies:

\[
P(E/H) > P(E) \rightarrow P(H/E) > P(H)
\]

This implies that if the hypothesis makes the evidence more likely, the evidence makes the hypothesis more likely. In other words, probabilistic dependence is bi-directional.

For example, if learning that someone eats well raises the probability that he is healthy, learning that he is healthy (and learning nothing else) raises the probability that he eats well. Although taking the converse of \textit{logical implication} does not preserve truth ((A implies B) does not imply (B implies A)), taking the converse of “A raises the probability of B” does preserve truth, in the absence of additional information justifying an overriding cross-induction. (See Chapter 2)

Bayes’ expositor Richard Price, who discovered Bayes’ Essay, and wrote the introduction to it, explains that Bayes’ “design” was “to find out a method by which we might judge concerning the probability that an event has to happen” given that “it has happened a certain number of times, and failed a certain other number of times.”

\textsuperscript{64} This statement trades on the vagueness of “vast”.
To justify such a probability judgment would solve the problem of induction, which appears to be a major part of Bayes’ aim. Price explains the relationship between Bayes’ project and the problem of induction, writing that Bayes aimed to “secure a sure foundation for all our reasonings concerning past facts, and what is likely to be hereafter.” (Price, Introduction to Bayes’ Essay, in Bayes 1763, pg. 1) He further explains:

The purpose I mean is, to shew what reason we have for believing that there are in the constitution of things fixt laws according to which things happen... It will be easy to see that the converse problem solved in [Bayes’] essay... shews us, with distinctness and precision, in every case of any particular order or recurrency of events, what reason there is to think that such recurrency or order is derived from stable causes or regulations in nature, and not from any irregularities of chance. (Dale 2003, 270-271)

If the existence of stable causes would increase the probability of observing vast regularities in the world, observing vast regularities would increase the probability of the existence of stable causes. The “converse problem” referred to by Price is the problem of deriving probability from observed relative frequency. A hypothesis that makes our observations of regularity more likely is itself made more likely by these observations. The hypothesis in question is that the regularities in nature are “derived from stable causes,” or that there are “in the constitution of things fixt laws.” The words “stable” and “fixt” imply transcendence of the modality (time, status as observed, etc.). Since fixed laws would make the observations of regularity more likely, the observations of regularity make fixed laws more likely. Because these laws are fixed in place, i.e. time-transcendent, they are able to ground predictions about the future or unobserved cases.

The argument is strongest if the hypothesis of stable causes is the only
hypothesis that would increase the probability of observing extensive regularities. Indeed, given a broad interpretation of “stable causes,” this hypothesis is the only such hypothesis; these “stable causes” or “fixt laws” may be defined as whatever it is that in fact increases the probability of the patterns we experience, making our experiences of persistent regularities probabilistically plausible, i.e. not colossally coincidental. Indeed, it seems that this open-ended interpretation is what Price has in mind. Considering “a die of whose number of sides and constitution we know nothing,” he writes:

...if in [the second] trial the supposed solid turns again the same side, there will arise the probability of three to one that it has more of that sort of sides than of all others; or (which comes to the same) that there is somewhat in its constitution disposing it to turn that side oftenest: And this probability will increase, in the manner already explained, with the number of times in which that side has been thrown without failing.65 (Price, in Bayes 1763, pg. 17)

Dispositions and causes “come to the same” as probability. Having reason to believe that there is something in the constitution of things disposing trials to turn out a certain way gives us reason to believe that it will probably turn out on the next trial the way it is thus concluded to be disposed. This disposition is, somehow, unlikely to disappear through the passage of time, at any particular time; it is “fixt” deeply into the “constitution” of things. But we need not make any detailed commitment as to its nature; we can understand it just as a probability.66

Bayes’ argument relies on the principle of indifference. For example, Price writes that “in the case of an event concerning the probability of which we absolutely

65 The quotation begins with an explanation of the claim that there is no probability assignment until the first two trials are complete.
66 However, Richard Price inferred much more, including, like Foster, the existence of God. See Price, Introduction, in Bayes 1763, pg. 2.
know nothing antecedently to any trials made concerning it... I have no reason to think
that, in a certain number of trials, it should rather happen any one possible number of
times than another.” (Price, in Scholium, Bayes 1763, pg. 11) Similarly, Bayes writes,
“Suppose the square table or plane ABCD to be so made and levelled, that if either of
the balls O or W be thrown upon it, there shall be the same probability that it rests
upon any one equal part of the plane as another, and that it must necessarily rest
somewhere upon it.” (Bayes 1763, Section 2, pg. 8.) This assumption of equal
probability later became a source of concern for many philosophers skeptical of the
solution.

3.3 Laplace
Laplace makes an argument similar to that of Bayes, writing, “In general, one finds in
this way that the constant and unknown causes that favour simple events which are
judged equally possible, always increase the probability of the recurrence of one and
the same simple event.” (Laplace, 35) Laplace, like Bayes, assumes the principle of
indifference: we begin with a judgment that the different possible “simple events” are
equally likely. We then find that our initial judgment needs to be adjusted as some
simple events are found to occur much more often than others. We infer that there are
unknown causes responsible for some simple events occurring more often than others,
which though initially judged “equally possible” (or equally likely), are now inferred
to be more likely to recur.

Thus we begin with flexible, uniform priors, and as experiences accumulate,
our choice of uniform priors plays a decreasingly central role. The Sun could just as
well set in the east and rise in the West, for all we know *a priori*. However, once we
have experienced it repeatedly rising in the East, we infer that there is something
cause it to rise in the East each time, increasing the probability that it will happen
again. Echoing Bayes and Price, Laplace writes, “When a simple event... has been
repeated a large number of times, the possibilities [probabilities] of the simple events
that maximize the probability of the observations, are those that are shown by
observation to be the most likely.” (Laplace, 38) Whatever probabilities would be
most likely to lead to the observations, are most likely.

Laplace recalls Bernoulli’s law of large numbers. He writes:

The probability that the ratio of the number of white balls drawn to the total number of
balls drawn does not differ from the ratio of the number of white balls to the total
number of balls in the urn by more than a given amount, tends to certainty as the
number of events keeps on increasing. (Laplace, 36)

We can draw an analogy to Hume’s problem of induction by making some
substitutions. Let “white balls” = cases in which induction is successful; let “balls
drawn” = cases observed; let “total number of balls in the urn” = total number of
cases: The probability that the ratio of the number of observed cases in which
induction is successful to the total number of observed cases does not differ from the
ratio of the number of cases in which induction is successful to the total number of
cases by more than a given amount, tends to certainty as the number of events
increases. In other words, the percentage of cases in which induction is successful in
our observations (of which there are many) is probably close to the percentage of
cases in which induction is successful overall. This gives a high probability to
induction continuing to be successful.
3.4 D.C. Williams

D.C. Williams makes basically the same argument. He explains the reasoning behind the law of large numbers solution to Hume’s problem: “any population will probably and approximately match, in any statistical respect, any of its samples.” (Williams, 79) In the case of singular inference, if the population is the observed plus the next unobserved case, both the observed cases and the unobserved case are samples of the same population, and are therefore likely to resemble each other. Williams defines the problem of induction “as the problem of showing that it is highly probable that a given sample… matches in statistical composition its population,” and solves the problem “by showing that the given sample is necessarily a member of a class of possible samples… of which the great majority do match” the sample. (Williams, 21) Since most large samples have frequencies similar to those of the population as a whole, it is most likely that the large sample that we have experienced is among this majority.

Campbell & Franklin explain Williams’s argument:

As most of the subsets in the hyper-population match the population, the chances are high that the subset we pull out will match the population. The situation is, in essence, no different from the situation where we pull a ball out of a barrel, where most of the balls are red. In that case, the odds are high that we will get a red ball; in induction, the odds are high that we will get a matching or ‘representative’ subset, because most of the subsets are representative. (Campbell & Franklin, 2004)

Induction is likely to be successful, for the same reason that we are likely to pick a ball out of a barrel which has the features of the majority. The future will probably resemble the past on the same grounds that polls are likely to reflect trends in populations beyond their samples.
3.5 De Finetti and Mackie: Alternatives to the Principle of Indifference

Like Bayes and Laplace, Williams relies on the principle of indifference, saying it is “an innocent truism” “that any one sample... is antecedently as likely to be selected as any other.” (Williams, 99) In contrast, many attempts have been made to justify induction by relying on weaker forms of an indifference principle. For example, “The successive attempts of Bayes, Johnson, and de Finetti to solve the problem of induction are marked by the invocation of progressively weaker symmetry assumptions.” (Zabell, 12) J.L. Mackie’s tolerance principle can also be regarded as an example of this. (See below) Bruno de Finetti’s notion of exchangeability is weaker in some respects than standard principles of indifference, and can replace it in an argument for induction:

By means of exchangeability it is possible to give an intrinsic characterization of inductive inference, because the probability of the various outcomes of future draws given the outcomes of the past ones is directly derived from the condition of exchangeability according to which all the permutations of the possible outcomes are given the same probability. (De Finetti 2008, 78)

Brian Skyrms explains the role of exchangeability here in grounding inductive reasoning, from probability in observed trials to probability in unobserved trials: “The probability of a property is just the probability that that property is exhibited on any trial-any trial, for by the hypothesis of exchangeability that probability is the same for any trial.” (Skyrms 1980, 8) We infer the probability of events from their observed frequency, within a margin of error. From the assumption of exchangeability we infer that the probability of an event is equal at all times, past and future. From the fact that

67 Here the similarity between the objection to the principle of indifference and the objection from the possibility of selection bias is apparent.
it has continued with such frequency in the past, we infer that the probability is high, for both the past and future.

A common objection is that an assumption of exchangeability ignores or rules out the possibility of relevant information about the order of our observations. (See Good, p. 21) For example, Peter Millican writes:

Suppose that the coin tossing machine were to produce as output:
000000000000000000000000000000000
Surely we would strongly suspect that things had changed... The very fact that we would in such a case see ordering as a relevant consideration again casts serious doubt on whether we would ever in practice consider a sequence of events to be literally exchangeable, or even if we did so initially, that we would then persist in Bayesian conditioning on that basis in the teeth of such evidence.” (Millican 1996, p. 163-164)

Zabell responds, “An appropriate generalization of exchangeability that takes such order information into account is the concept of Markov exchangeability.” (Zabell, 12) Markov exchangeability is the property such that all strings with the same initial letter and same transition count have the same probability. But we can question the grounds for accepting an assumption of Markov exchangeability, and Millican is right to challenge the justification of exchangeability, Markov exchangeability, or whatever variant might turn out to best fit our most confident beliefs about induction.

Millican misplaces the objection, however, writing, “If De Finetti depends upon background assumptions of real physical propensities... to explain the exchangeable patterns of beliefs from which his theory takes off, then as far as the justification of induction is concerned, we are no further forward.” (Millican 1996, p 165) De Finetti need not rely on such an assumption: assumptions of probabilistic properties are sometimes present even in the absence of assumptions about physical
objects, such as in mathematics.

This point is both common, and commonly overlooked. Keynes writes that when a formula “has been verified in every case in which verification is not excessively laborious... we feel that this is some reason for accepting it... Yet there can be no reference here to the uniformity of nature or physical causation.” (Keynes, 242) Similarly, Franklin writes that “the view that the problem of induction should be solved in terms of natural laws,” does not take account of the fact that induction works in mathematics “just as well as in natural science.” (Franklin, 1987) Induction with respect to mathematical patterns can be grounded in an assumption of exchangeability no less than induction with respect to the natural world, so there is no prima facie reason to think that an assumption of exchangeability would depend on “a background assumption of real physical propensities.” Nonetheless, the criticism that De Finetti does not justify the assumption of exchangeability, though not a problem for an “intrinsic characterization of induction,” as De Finetti refers to his project, would be a problem for an attempt to solve the problem of induction.

The principle that epistemic probability supervenes on psychological structure plays the same role in grounding induction as exchangeability but is independently intuitively motivated. In Groundhog Day, given that my experience today is the same as yesterday, my epistemic probabilities are the same. Thus we can use the supervenience principle, within this framework, to achieve the same result as exchangeability: equalizing present and past probabilities.

Mackie offers another alternative to the principle of indifference with his “principle of tolerance,” or non-zero priors for probabilistic dependence:
...it is arbitrary to lay it down dogmatically *a priori* that the distribution of uniformity must be purely random. If, instead, we allow from the start that there is some better-than-zero probability of some not purely random pattern—as opposed to purely local appearances of order which really result from chance—then the inverse argument can work in the modest but reassuring way outlined above. (Mackie, 1979, 175)

Thus we can argue that observed data is relevant to the probability of unobserved data using less than the principle of indifference, namely a principle of tolerance, to which the standard objection to the principle of indifference does not seem to apply.

As long as it is not *a priori* extremely unlikely that probabilities over time are interdependent, observations of vast and stable regularity in the past will increase the probability that they are, and therefore increase the probability that regularity will continue for some further length of time. If the problem of induction is reduced to the problem of justifying the assumption of non-zero priors for probabilistic dependence, the burden of argument shifts. It is *prima facie* puzzling what justifies our belief that the future is likely to resemble the past, but not what justifies our belief that this likeliness is an epistemic possibility with probability greater than 0. This is a different question. Our understanding of the world does not seem to rule out an epistemic or metaphysical probabilistic connection/interdependence between observations or events at different times. It is a reasonable assumption, if even an assumption at all, that induction *might* be justified (with greater than 0 probability). The burden is on the skeptic to say why we should be worried about such an extreme, *ad hoc* proposal as the denial of a greater than zero prior probability for probabilistic dependence. Hume would not have generated much interest, including his own, if he had put the challenge to induction in this anachronistic way. The worry about the ground of non-zero priors
for probabilistic dependence of events is the last refuge of a Humean radical skeptic painted into a corner.

3.6 Blackburn

There is a way to state the problem of induction which suggests the solution. Supposing that we cannot provide a reason to believe that the future will probably resemble the past, we can ask how, in the absence of a reason, our past inductions could have been so successful. Surely it wasn’t just a colossal coincidence that time after time, we guessed correctly that the Sun would rise despite there being no high probability of our guesses being correct. So, we must have had a reason, a justification, even if we cannot identify it. The alternative, that we have been colossally lucky, is implausible.

This response blocks the circularity objection from the start. If induction was justified in the past, it must be justified in the present and future. The same argument, with the same premises and conclusion, cannot be justified at one time and not another. Unless it is a colossal coincidence that it has worked so well, which is a priori unlikely, there is a reason it has worked so well, something about inductive reasoning that makes it inherently likely to work.

Simon Blackburn makes an argument that can be understood in this way. In reference to an example of Roy Harrod, he writes:

We are told that [the sceptic] knew that a regularity had lasted ten years, and used this fact to justify not taking the average success of one-tenth continuance predictions as relevant; but counted the contrary prediction as grounded just as well or badly... since in general use of this form of argument leads to the view that false predictions are as justifiable as true ones, and nothing further is known about this case, no weight should be attached to the sceptic’s manoeuvre. (Blackburn, 149-150)
“Why does induction work?” is both the problem and the solution. Unless the problem of induction has a solution, this question has no answer. But in Blackburn’s view, induction worked in the past because it was rationally justified, just as deduction worked because it was justified. Otherwise, it would be an implausible colossal coincidence.

When one person consistently makes more accurate predictions than another, it seems likely that there is a reason. Even if we cannot imagine the specific mechanism, it seems likely that there is one. For example, they may have additional knowledge or superior powers of inference.

The key point is that the charge of circularity is avoided by relying on the timeless nature of something, namely rationality. If induction (or mathematics, or any form of reasoning) is rational at one time, it is rational at all times; the rules of rationality are timeless/tenseless, not subject to change. I think that any solution to the problem of induction must rest on an inference to something which is timeless in this sense if it is to avoid circularity.

Unlike many authors who offer a solution involving similar reasoning, including Harrod (whose solution Blackburn’s grows out of), Blackburn’s solution relies on the past regularity of inductive inferences tending to be successful, not directly on past regularities in nature more generally. For Blackburn, the fact that it works better than its alternatives suggests its rationality. D.C. Williams explains the point: “In describing the logical cogency of induction, we by the same act explain why... it commonly works.” (Williams, 164) The idea can be illustrated by an analogy
to the fact noted above that mathematical theorems, which are later proven, are often first suspected on the basis of (non-mathematical) induction, namely the observation that in case after case in which it is tested the theorem holds. The fact that the theorem works holds with such consistency that it is at least somewhat likely that there is a proof. Similarly, since induction works, we infer that it is likely to have a justification – perhaps even this very same argument.\(^6\) Whereas James’s pragmatism defines truth as success, this account infers probable truth from success.

However, it would be “anthropocentric” to hold that induction is only justified in cases in which it has actually been used and found frequently successful. Even if there were no conscious beings before now, as long as we know that the Sun has always risen persistently in the past, we know that it will probably continue to rise. It is enough, therefore, to know that if induction had been used, it would have been successful. The premise of extensive past regularity is sufficient even if no inductive reasoning took place in the past. Regularity or patternedness is sufficient for rational induction - we do not specifically require regularity of inductive success. Nonetheless, Blackburn’s solution seems correct, if interpreted in terms of success for hypothetical inductive reasoners.

3.7 Foster, Armstrong, and Bonjour

Foster, Armstrong and Bonjour justify induction in terms of inference to the best

---

\(^6\) Hume seems to recognize the need to explain the success of our inductive inferences, appealing to a “pre-established harmony”: “Here, then, is a kind of pre-established harmony between the course of nature and the succession of our ideas; and though the powers and forces, by which the former is governed, be wholly unknown to us; yet our thoughts and conceptions have still, we find, gone on in the same train with the other works of nature.” (Hume, E5.2)
Foster outlines his “nomological-explanatory,” inference to the best explanation, solution to the problem of induction:

The first of these two steps is an inference to the best (the most plausible) explanation. What the explanation is advanced to explain is the occurrence of the hitherto exemplified regularity… And this regularity calls for explanation because it is too extensive to be deemed coincidental – deemed to be something that has occurred for no reason… This explanation involves the postulation of some law or set of laws of nature, sometimes precisely specified, sometimes not… The second step of inference is a deduction from the explanation. (Foster, 58)

Foster invokes the reasoning that there must be an explanation of past regularity, the presence of which enables us to avoid accepting the existence of a colossal coincidence.

The statement that “regularity calls for explanation because it is too extensive to be deemed coincidental,” suggests that by an “explanation,” Foster means a description of past regularity which lessens its coincidence. Similarly, Bonjour writes, “In a situation in which a standard inductive premise obtains, it is highly likely that there is some explanation (other than mere coincidence or chance) for the convergence and constancy of the observed proportion.” (Bonjour, 208) Like Foster, Bonjour here seems to implicitly define explanation as the reduction or elimination of coincidence. Therefore, we can simplify these theories by appealing directly to lesser coincidence rather than inference to the best explanation, avoiding some of the ambiguities of “best” and “explanation”.

Foster encounters the Humean circularity objection in the form of an objection to his inference to the best explanation justification of induction, namely the

---

69 The indirectness/complexity of inference to the best explanation is manifested in the two steps of the “nomological-explanatory” model, in which induction is justified through a combination of abductive reasoning to explanans (law) followed by deductive reasoning to prediction.
possibility of a time-restricted law. (Foster, 70) He argues that the best explanation of past regularity is that there are laws of nature. (Foster, 58) He must then choose between the competing laws, “It is a law that bodies always behave gravitationally,” and “It is a law that, at all times before T, bodies behave gravitationally; and there is no stronger gravitational law.” (Foster, 70) The Humean objector argues that the only way to justify favoring the first over the second law is by assuming that the laws are likely to be uniform, therefore circularly assuming a form of the principle of uniformity.

Foster’s response is unconvincing. According to Foster, the time-restricted law explanation is not the best explanation, because it is “mysterious” and “puzzling.” (Foster, 71) He writes, “Why should a certain moment have this unique significance in the structure of the universe, that bodies are gravitationally constrained in the period up to it, but not thereafter?” (Foster, 71) He writes that “irrespective of what happens in practice, there is something very odd in the abstract about a situation in which events keep on conforming to a common pattern by chance—without anything to prevent or discourage deviance.” (Foster, 62) By “in the abstract,” Foster seems to mean “a priori”. He argues, “For it seems to me that a law whose scope is restricted to some particular period is more mysterious, inherently more puzzling, than one which is temporally universal.” (Foster, 71) He says that an unrestricted law is “decisively more plausible than any alternative explanation.” (Foster, 71) But Foster does not provide much elaboration, and it is not clear that his response amounts to more than the assertion that induction is likely to yield correct results, on the grounds that a world in which it does appears, intuitively, more plausible. It seems that we do not
have a right to appeal to bare intuition at this step. If we can help ourselves to intuition at this step, why not skip inference to the best explanation and the inference to laws altogether and say from the beginning that a world in which the future doesn’t resemble the past would be mysterious or puzzling, and conclude immediately that the future will resemble the past?

Foster writes:

Thus if someone were seriously to propose [the time-restricted law] as the correct account, our response would be to ask why the relevant law should be time-discriminatory in that way. Why should a certain moment have this unique significance in the structure of the universe, that bodies are gravitationally constrained in the period up to it, but not thereafter? (Foster, 71)

The Humean objector can simply reply, “Why not?” Why shouldn’t the law be time-discriminatory? Perhaps Foster is implicitly appealing to a principle of sufficient reason, like that of Leibniz – there could be no reason for one moment to be chosen over another as the time when the laws change. But this interpretation is speculative, and the principle of sufficient reason, especially in this application, is dubious.

More importantly, we should be wary of accepting the heavy empirical consequence of Foster’s theory that no moment will have this “unique significance,” i.e. that the regularities of nature will never immediately cease or change. This result is too strong. All we need to conclude, and all we are justified in concluding, I claim, is that if there is such a moment in the story of the Universe, it is unlikely to be now.

Foster continues, regarding the question he raises of why the law should be time-restricted:

Any answer we could receive would only serve to show that the nomological situation was not as suggested—that the change in the constraints on gravitational behavior was
to be ultimately explained in terms of time-impartial laws and a difference, relevant to the operation of these laws, in the conditions which obtain in the two periods. (Foster, 71)

Here Foster claims, correctly, that time-partial laws must be based in time-impartial laws. But his argument is unclear; it’s not clear why Foster thinks there is likely to be an answer to the “why” question.

Because these why-questions seem pertinent and yet are *ex hypothesi* unanswerable, we are left feeling that, as hypothesized, nature would be inherently puzzling, and would preclude an explanation of our empirical data which was both correct and, from the standpoint of our rational concerns, fully satisfactory. (Foster, 71)

He writes that the time-impartial law hypothesis is “decisively superior” because it “dispels one mystery without creating another: it dispels the mystery of unexplained regularity without creating the mystery of capricious necessity.” (71) This argument seems to (but, I claim, does not) rely on a false analogy: Unexplained regularity (i.e. coincidence) is “mysterious” in a special way which allows us to infer that it is *a priori* unlikely to occur in any given instance. That coincidence is unlikely is analytic; if a co-occurrence were likely, it would not be coincidental. But Foster has not shown that the mysterious, “capricious necessity” of a time-restricted law is unlikely in the same way. Still, Foster is right that only a time-impartial law account “dispels one mystery” without “creating another,” if by “mystery” we mean coincidence. (Foster, 71)

Foster adds that the idea that the concept of ‘law’ is unamenable to time-restrictions. He writes, “unlike the concept of behavior, the concept of a law of nature has some notion of generality built into it.” (Foster, 72) But this just raises doubt as to whether the best explanation of observed regularity is a law in this sense, or instead a
quasi-law, with all of the properties of a law except generality. Foster has not shown that the concept cannot be reshaped in this way.

Bonjour offers a more persuasive response to the time-restricted law objection, which I will describe and expand on. He writes that we need a notion “involving by its very nature a substantial propensity to persist into the future” in order to explain the regularity observed, because “anything less than this will not really explain why the inductive evidence occurred in the first place.” (Bonjour, 215) Bonjour defines “the best explanation,” correctly in my opinion, as “the most likely to be true.” (Bonjour, 212) Suppose that we accept the explanation that there is a time-restricted law, or time-restricted high probability, that bodies behave gravitationally, restricted exactly to the past. The reason for accepting any such explanation of the observed regularity is to avoid the improbable conclusion that the observed regularities of nature constitute a colossal coincidence. In the absence of this high probability, the abundance of known cases of bodies behaving gravitationally would indeed constitute a colossal coincidence. However, if we suppose that the probability in question is time-restricted, then in the absence of a second-order probability grounding its persistence, its own persistence would be equally coincidental. How do we know, then, that the second-order probability isn’t restricted to the past? If it is, we can ask the same question of how, in this case, it could have persisted over time in the past, unless there was a higher-order probability for it to persist. We can follow this chain of reasoning up the ladder, _ad infinitum_, until a time-impartial higher-order probability is eventually reached which the chain of time-restricted probabilities of time-restricted probabilities rests upon, or colossal coincidence is, absurdly, admitted. This response parallels my
reply to the circularity objection in general. (See Ch. 4)

A time-restricted probability account, in the absence of the posit of a higher-order time-impartial probability, has no “explanatory” force at all. It is not an alternative to a colossal coincidence account; it is a colossal coincidence account in which the coincidence is shifted to the persistence of the time-restricted probability. Only a probability which is not subject to change with time requires no higher-order probability in order for its persistence for any significant length of time not to be colossally coincidental.

Foster correctly asserts that the time-restricted law account but not the time-impartial law account simply “dispels one mystery” while simultaneously “creating another”. The “mystery” in question is the implicit supposition of colossal coincidence. Hence the response does not rely on further assumptions about which possibilities are mysterious and how this might lower their probabilities, as Foster’s response seems to suggest. Positing time-restricted laws, just like positing the absence of laws altogether, leaves us with a colossal coincidence account of our experience.

*Timeless* high probability for vast regularity in progress to continue is not merely the “best” explanation of observed vast regularity, it is the only explanation in the sense relevant to Hume’s central problem of induction, as the only description that lessens coincidence and is thereby confirmed.

This reply has the additional advantage over Foster’s that it does not rigidly claim that it is unlikely that the laws of the Universe will *never* change, claiming only that it is unlikely that the laws will change now, or at any other *particular* time in the future up until which it will have persisted. We are not warranted *a priori* in supposing
that the laws of physics will *never* change. This question lies outside of philosophy. Bonjour agrees, writing that the propensity that we infer as a result of observing vast regularity in nature “need not... be so strong as to rule out any possibility that ‘the course of nature might change,’ but it must be sufficient to make such a change seriously unlikely.” (Bonjour, 215) This protects us against having to rule out a priori intuitively open empirical possibilities.

In contrast, we are warranted in supposing, a priori, that the high probability that vast regularities in progress will persist into the next case can never change. Foster writes, “because it is naturally necessary that bodies behave gravitationally, it logically follows that they always do.” (Foster, 69-70) But we are not warranted in asserting that it is naturally necessary that bodies behave gravitationally - we should conclude instead only that where bodies have been observed to behave in this way, they are likely to continue to do so for some further set of observations. Even if we expect the laws of nature to change at some time in the future, so that induction will at that time fail to produce successful predictions, it is ad hoc to think that the time is now, and unlikely to be correct. Analogously, we know in advance that a TV will stop working eventually, but it is unlikely to stop working at any particular time.70

Armstrong’s account is very similar to Foster’s. Armstrong writes:

> We need an explanation of the rationality of induction... My own explanation is this. The sort of observational evidence which we have makes it rational to postulate laws which underlie, and are in some sense distinct from, the observational evidence. The inference to the laws is a case of inference to the best explanation... If the inferred laws exist, then, of course, they entail conditional predictions about the unobserved. (Armstrong, 52-53)

---

70 Example suggested by John Perry.
These laws he understands as “a connection between universals,” which “can provide an explanation of an observed regularity in a way that postulating a Humean uniformity cannot.” (Armstrong, 104) It is a side issue, with respect to solving the problem of induction, what the correct explanation of past regularity is, i.e. which alternative to the Regularity theory is correct. Any explanation in the relevant sense is sufficient to justify induction. I am agnostic about Armstrong’s theory of universals, but it is not necessary to fill out such details about the nature of the explanation inferred for the purposes of solving the problem of induction. We do not always need to know exactly what it is we are accepting in order to have grounds for accepting it.

The significance of the difference between a less committal view of the nature of the “explanation” and Armstrong’s reliance on a more specific view is illustrated by Armstrong’s limited resources for responding to the objection “that laws construed as irreducible relations between universals” might not “constitute the best explanation of the observed regularity of the world,” given that there are “other possible explanations of the regularity of the world.” (Armstrong, 105) Armstrong’s over-specificity forces him to address the question of how one can know that his particular explanation is best. Perhaps Armstrong’s explanation is only the best theory that anyone has come up with yet – not the best possible. But there is no need to arbitrate among every possible explanation; all explanations, qua explanation in the relevant sense, are sufficient to ground the rationality of induction. In contrast, Armstrong’s response seems weak:

The answer to doubts such as these is simply to admit them. There can be no guarantee that the explanatory scheme which we favour is in fact the best explanatory scheme. All that can be done is to spell out all proposed schemes as fully as possible, and try to see which scheme fits the apparent facts best. (Armstrong, 105)
Expanding on what he means by “best”, he says that a good explanation should “genuinely unify” and “be genuinely informative.” (Armstrong, 105) Armstrong writes, “The only answer to that is to challenge the questioner to find a better explanation.” (Armstrong, 53) But there is more to be said. It doesn’t matter which explanation is correct; we need not privilege one in particular. Any explanation in the relevant sense of the term, i.e. any hypothesis that is an alternative to colossal coincidence, is sufficient to justify induction. Whatever the best or correct explanation might be, it justifies induction. We can concede that regularity might be explained by something other than necessary laws, but any explanation must be likely to persist in order to be an explanation, in the relevant sense of lessening the unlikelihood of the persistence of something else.

Armstrong considers a second, related objection: “More fundamentally, [the objector] might ask what reason there is to think that the observed regularities of the world have any explanation at all.” (Armstrong, 105) Again Armstrong’s response seems weak: “It is surely a great advance in the battle against the sceptic to be able to say ‘This is a, perhaps the, rational explanation of the phenomena, although it is logically possible that the phenomena have no rational explanation.’” (Armstrong, 106) Armstrong does not elaborate on how it is an advance. His response here resembles Reichenbach’s claim that induction is likely to succeed if any method is, and is equally unreassuring.

At this point, like Strawson and Foster, Armstrong appeals to meaning. He says, “We can finally appeal to the meanings of words, the appeal which Strawson and others made too soon. To infer to the best explanation is part of what it is to be
rational.” (Armstrong, 59) But it seems that Armstrong does not hold out much longer than Strawson. Armstrong offers no proof that the future is likely to resemble the past, giving us at best a prediction-grounding explanation which is “perhaps” correct, and claiming that this is all that rationality requires. But if we can justify only the statement that perhaps the Sun will rise, not that it probably will, we have not solved Hume’s problem.

Like Foster, Armstrong describes and then rejects the sampling version of the solution. He writes, “A purely mathematical argument can be mounted. It is a necessary truth of arithmetic that a high proportion of large samples of a population match the distributions in the population itself.” (Armstrong, 57) He argues that Goodman’s new riddle of induction “is a conclusive reason for thinking that the principles of logical probability cannot, by themselves, solve the Regularity theorists’ problem.” (Armstrong, 57)\(^7\) In order to solve the riddle, he invokes his account of universals. It is unclear, however, why he does not advocate the sampling argument in conjunction with his account of universals. Moreover, Goodman’s riddle should not be mistaken for the problem of induction. (See Ch 1.)

But Armstrong correctly emphasizes the fact that “no Regularity theorist, whether or not he is prepared to call his regularities ‘laws’, can escape inductive scepticism.” (Armstrong, 5) The Regularity theorist holds that “laws of nature are nothing but Humean uniformities”; a law, on this view, is just the set of its actual cases; causal facts supervene on what actually happens. (Armstrong, 52) Armstrong explains why an alternative to the Regularity theory must be supposed by any

\(^7\) Thus Fales’ objection to Foster’s solution to Goodman’s riddle that Foster must, supposedly, assume realism about Universals, seems more appropriately targeted at Armstrong. See Ch. 1 above.
adequate solution to the problem of induction: “If the law is simply the conjunction of the observed and the unobserved cases, no appeal to the law can have any value. The inference will be a straight inference from the observed to the unobserved.”

(Armstrong, 104) A law in the sense that is consistent with the Regularity theory, determined by actual physical occurrences rather than counterfactuals, does not explain the regularity in the world; it is just an assertion of this regularity (and, perhaps, additional regularity).

Similarly, Bonjour writes, “the assertion of a Humean constant conjunction amounts to just a restatement and generalization of the standard inductive evidence, but has no real capacity to explain the occurrence of that evidence.” (Bonjour, 215) What we need instead, says Bonjour, is a notion of “objective regularity” conceived “as involving by its very nature a substantial propensity to persist into the future.” (Bonjour, 215) The only relevantly explanatory fact would be one that inherently, “by its very nature,” independently of accidental properties like place in time, renders conjunctions such as those that have been observed likely. Where Armstrong writes of laws of nature as relations among universals, Bonjour writes less committally of propensities. However, it seems to me that even this assumes too much. “Propensity” suggests a physical world responsible for our experiences and epistemic probabilities, more than needs to be assumed by a solution to the problem of induction.

But it is true, as Bonjour and Armstrong agree, that our propensities, laws or probabilities must amount to more than mere regularities. Bonjour writes, “The justification for conceiving the regularity in this way is that anything less than this will not really explain why the inductive evidence occurred in the first place.” (Bonjour,
Whatever the correct explanation, if any, it must have an inherent likeliness to persist, and can therefore be used to rationally ground induction. On pain of colossal coincidence, it is probable that some such correct explanation of the relevant kind exists, whether or not it is among our current hypotheses.

In contrast to Foster and Armstrong, Bonjour seems to imply that the inference to the best explanation argument is equivalent to the statistical sampling argument: “The best explanation, that is, the most likely to be true, for the truth of a standard inductive premise is the straight inductive explanation, namely that the observed proportion \( m/n \) accurately reflects... a corresponding objective regularity in the world.” (Bonjour, 212) Unlike Foster and Armstrong, Bonjour here correctly identifies “best” in “best explanation” with “most likely,” as opposed to identifying it with least “mysterious,” or rational by definition. Bonjour qualifies this by bracketing the anthropic “possibility that observation itself affects the proportion of \( As \) that are \( Bs \).” (Bonjour, 212)

3.8 Chapter conclusion

Expressing the solution in terms of coincidence and likeliness helps bring the various formulations of the solution together, enables us to avoid disputes about the meaning of “best explanation” in science and ontological issues about the world behind the phenomena, and allows us to see that the time-restricted law objection does not present a genuine alternative to statistical absurdity.
Chapter 4. Solution, Objection, and Replies

4.1 Solution

For variety, let us rearrange the argument’s logical structure, speak in terms of epistemic probability instead of coincidence, and express a suppressed premise (iii):

Premise (i) Either vast regularities in progress are time-impartially epistemically likely to continue to the next case, or it is extremely unlikely that we would have observed vast regularity.
Premise (ii) We have observed vast regularity.
Premise (iii) If something is extremely unlikely, it is more likely that its negation is true, other things being equal.

--------

Conclusion (c1) It is more likely that vast regularities in progress are time-impartially epistemically likely to continue to the next case, other things being equal.
Conclusion (c2) In Groundhog Day (where other things are equal because there is no additional cross-inductive information available), given that the Sun has risen with vast regularity in the past, it is more likely that it will rise again in the future.

The argument is an instance of inference from the low probability of a coincidence to the higher probability of the disjunction of alternatives. The observed vast regularities constitute a colossal coincidence unless vast regularities in progress are time-impartially epistemically likely to continue. Therefore, vast regularities in progress are, probably, time-impartially likely to continue.

The underlying idea behind this argument is ancient, long preceding Hume’s question (to which it supplies a pre-emptive answer). Cicero writes, “...nature’s punctual regularity... is due to nature’s forces.” (Cicero, 313) From the observed patterns, we infer forces. These forces are the reason our observations of nature fall into the patterns they do. There must be some reason, and forces, connections, and the like are just defined as that reason.

Hume claims that observed patterns only reveal, at best, forces in observed
instances, not in unobserved instances. But when we reflect on the rational motivation for accepting the existence of forces in the past, in Cicero and others, we see that its epistemic/explanatory purpose is only fulfilled if the inference holds for unobserved cases as well as observed cases. The inference to forces is motivated by a theoretical/quasi-mathematical preference for coincidence reduction. If forces aren’t tenselessly likely to continue after they have long persisted, no coincidence reduction is achieved by positing them. In order to be fair to the spirit of Cicero’s inference, we must suppose that what he means by force carries with it a time-transcendent likeliness for its own continuance, i.e. there is a high probability, regardless of when, for forces that have long persisted to continue for some further length of time.

Were there no such high probability, the fact that the forces have persisted would itself be just as unlikely as our observations of regularity would be in their absence. The appeal to powers would therefore be explanatorily powerless, failing to “explain” in the sense of coincidence-lessening. Thus Hume’s circularity challenge ignores the argument for the past existence of forces (while claiming to accept their past existence for the sake of argument): forces that are not likely to persist at any time do not achieve the epistemic/explanatory aim for which they are intended, so there is no reason to believe in them. Foster writes, “We are justified in postulating a law of gravity... because it eliminates what would otherwise be an astonishing coincidence.” (Foster, 49) When Hume writes that it shall “be allow’d for a moment, that the production of one object by another in any one instance implies a power,” he allows

---

72 This principle of coincidence reduction may be the basis of Ockham’s Razor. Newton writes, “We are to admit no more causes of natural things than such as are both true and sufficient to explain their appearance.” (Newton, Book III, Rule 1, pg. 384) The simplest sufficient explanation is the least coincidental, because it would be more unlikely for a stronger hypothesis with more assumptions that aren’t independently motivated to be correct.
only the conclusion, not the reasoning behind it; namely, that a power (if and only if it is timelessly likely to persist) would lessen the coincidence of the observed regularities. (Hume, T1.3.6)

Observations of extensive patterns suggest causal or other probabilistic dependence. If we accept the basis for believing this, we must also accept that the dependence in question transcends the mode that separates the observed from the unobserved, such as time. Otherwise, the inference to dependence would be unmotivated, as the observed extensive patterns would be rendered no less improbable. The improbability of the extensive patterns would be transferred to the equally great improbability of the repeated persistence of this imagined non-mode-transcending dependence.73

Extensive patterns must have been likely to persist in the past, given that they persisted over many instances. This likeliness was either time-impartial or time-restricted. If time-restricted, the likeliness itself must have been likely to persist (a second-order likeliness), given that it persisted over many instances. The second-order likeliness is either time-impartial or time-restricted. If time-restricted, it must have been likely to persist (a third-order likeliness). Only a timeless likeliness allows escape from higher-order time-restricted likeliness of time-restricted likeliness ad infinitum, a building with no foundation, which never reduces the coincidence of our observations of vast regularity.

If something is subject to change with time, we can always ask why it holds constant at a certain time. If there is no answer, no reason why it exists at some times

---

73 To instead appeal to a yet higher-order mode-transcending dependence to explain the persistence of this non-mode-transcending dependence would admit as much as needed to solve the problem of induction, by guaranteeing that dependence is likely to persist at any time.
and not others, yet the trail of particular times at which it exists forms a pattern (such as being consecutive), this is a coincidence. Only something eternal, not subject to change with time, like a fact of mathematics or logic, escapes the endless “why” loop. For example, there is no question as to why the sum of 2+2 does not change over time. In contrast, a “law” that is no more likely to continue than to cease at each nanosecond is unlikely to hold for long. This is why a law in the sense allowed by the Regularity theory is insufficient for the kind of explanation that the observed regularities call for; consisting only of events at particular times, it is not time-impartial.

Coincidences are *a priori* unlikely in any given case. The greater the coincidence, the more unlikely it is to be true. Colossal coincidence descriptions of the world are for this reason extremely unlikely to be true. The only alternative to the possibility that our observations of vast regularity in nature are colossally coincidental would give us reason to believe that vast regularities are likely to continue, as a time-transcendent (timeless) fact. If existing extensive patterns are not timelessly likely to continue, extensive patterns would be extremely unlikely to arise, and colossally coincidental if ever they do arise. Given that extensive patterns have arisen, it is more likely that extensive patterns are timelessly likely to continue.

The world contains vast regularities (frequencies, patterns): nature exhibits stable frequencies; patterns persist; similar “causes” tend to have similar “effects”. Either this vast regularity is, at least for the most part, coincidental, or not. If not, this implies that existing vast regularity must be likely to continue, because otherwise it probably would not have come to be so vast in the first place; it would have ceased one of any number of times in the past at which it was unlikely to continue, and only
by colossal coincidence could this fail to occur. It is more likely that it is not a coincidence that it did not cease at any of those moments, because coincidences are, by definition, unlikely in any given case. Just as actual implies possible, frequent implies probable.\textsuperscript{74} From the high frequency of patterns continuing in the past, we infer high probability of such continuance in the past; from this, the principle that epistemic probability supervenes on psychological structure, and similarities among the present and past psychological states, we infer high probability of patterns continuing in the future. If patterns were epistemically likely to continue in the past, they remain likely, as there are no differences in the appearances between now and then that could support a difference in this probability.

This solution relies on a form of reasoning we can call inference to the only alternative to colossal coincidence (IOACC, a refinement of inference to the best explanation), or near-proof by near-contradiction.\textsuperscript{75} This involves reasoning by analogy to equate past probabilities with present probabilities: It must be the case that, from our present point of view of observing persistent regularities in our past, these regularities will probably continue, since otherwise it would be a colossal coincidence that they continued in case after case in the past after our point of view was the same as it is now.

Reasoning by analogy is most straightforward when the analogues are exact duplicates, as are all of the days in the past in the case of Groundhog Day. In

\textsuperscript{74} More precisely, epistemically frequent (frequent relative to what we know) implies epistemically probable. Metaphysically frequent does not strictly imply metaphysically probable, as metaphysically improbable events may, against all probability, occur frequently in a finite number of trials.

\textsuperscript{75} Proof by contradiction is included as the limit case of near-proof by near-contradiction, allowing us to adopt the latter as a generalization of an already independently accepted form of reasoning, rather than having to “invent” a new one from scratch.
Groundhog Day, all of my memories, perceptions, thoughts and beliefs have appeared exactly the same every day.\(^6\) Thus the probabilities of the same future perceptions occurring the same length of time in the future are also the same. Therefore, the fact that “the future” has turned out to resemble the past in our past experience gives us reason to believe that the future will likely resemble the past.

This inference to lesser coincidence solution may be thought of as an elaboration of the solution reported and seemingly refuted as circular by Russell. (See Ch. 1.) The argument is not circular, however, because epistemic probability supervenes on psychological structure. We can understand identical psychological structure, in Humean terms, as exact resemblance. Certain ideas resemble each other. If two people have exactly resembling complete sets of ideas, then, intuitively, the same epistemic probability norms apply to each of them, and they have corresponding reasons for believing corresponding things. Epistemic probability supervenes on the information intrinsically available, which in the case of Groundhog Day is the same each day.

Even in the messy conditions of the actual world, we can apply the same reasoning with a lesser degree of precision: things seem the same now as in the past, with respect to anything we think likely to be relevant to the question of whether the Sun is likely to rise on our respective tomorrow. If the Sun is unlikely to rise tomorrow, it was unlikely to have risen today, the previous day, and so on, and so this pattern would constitute a colossal coincidence. Therefore, it must have been likely to

\(^6\) John Perry asks, “How would we know this given?” A skillful wine taster can distinguish his experiences more finely. Perhaps what seems to us the same from day to day in Groundhog day is a function of an undeveloped ability to discern the details of our own experiences. But this is not Hume’s problem of induction. We can accept for the sake of argument, \textit{per impossible}, that our experiences are known to be identical and still raise Hume’s problem.
rise the next day each time, so it must be likely to rise tomorrow. Thus, the circularity objection cannot be raised here, because concerning what is likely, we have grounds for believing the cases to be similar, as things seem roughly the same and what is likely is determined by how things seem.

It must be true that in worlds exhibiting great regularity up until an arbitrarily observed time, it is likely that regularity will continue for at least some further length of time. Otherwise, we would probably not be in a world of such great regularity. Having experienced all of the regularity we have in the past, it would still be unlikely to continue for one second if this principle were not true; moreover, the probability that regularity would continue for yet another second, conditional on continuing for one second, would also be very low. For example, the probability that my hand would continue to exist for even a minute longer would be extremely small, and it would be a colossal coincidence if it happens. One minute later, my hand still here, I must either accept the premise that in worlds exhibiting vast regularity, regularity is likely to continue, or believe that a colossal coincidence has just taken place.

Formulating the argument in diachronic terms - rather than giving only a backwards-looking explanation of the regularities - corresponds to the scientific practice of assigning greater credence to theories that are initially put forward on grounds independent of the data that later confirm them, such as the confirmation of Einstein’s theory of general relativity by observations of bent starlight. (Had Einstein

---

77 Here I am applying probabilities to possible worlds, not analyzing probabilities in terms of proportions of possible worlds.

78 Again, I bracket the anthropic alternative that we would not be asking whether induction is likely to succeed in the future if not for the vast regularities in the world so far observed, as we would not be conscious inquisitors. Thus, according to this anthropic reasoning, it is not a coincidence that the vast regularities of the past have occurred, because the choice to ask the question involves selection bias in that it is only asked in worlds in which a colossal coincidence occurs. I’ll only remark that this anthropic reasoning seems to commit the theorist to the existence of a multiverse.
designed the theory of general relativity in order to explain these observations in hindsight, the evidence for the theory may have been less convincing.) If, at every second, it is no less likely that we live in a world where the regularity will immediately cease than that we live in a world where it will continue for some length of time, we must predict that the extensive regularities we have observed in nature will very soon cease, and if we wait a few seconds, we will surely find this prediction contradicted. If, on the other hand, we predict that the extensive regularities of the past will continue at least a little while longer, time will undoubtedly prove us right. Following Hempel’s duality thesis, we can then call this prediction’s predictans (i.e. the basis for the prediction) an explanation (explanans).

That the observed and unobserved facts or impressions are represented as separate does not mean that they are objectively separate, and our experience can provide evidence of their objective non-separateness. This may be or involve a kind of identity, or, as Mackie says, something beyond mere “four-dimensional machinery,” including a notion of “things as persisting through time and processes as projecting themselves in time.” (Mackie 1985, 175, quoted above) Or, perhaps it is mere epistemic probability.

As an analogy, imagine that we find a quarter on the street, which upon first inspection looks like an ordinary, fair quarter. We have a strong bias, before flipping the quarter, in favor of the belief that the quarter is fair, given that most and perhaps all quarters in our experience have been fair (or very nearly so). But if every time we flip the coin, it lands heads, even after thousands of flips, then we should believe that it is more likely to land heads on the next flip. It is probably not a coincidence that it
has landed heads so regularly, and so there is probably a reason why it has done so. Independent of speculation about what the reason is (e.g. an uneven weighting of the coin, a magnetic pull between the ground and the heads side of the coin, the will of an omnipotent being, or perhaps some circumstance we now find unimaginable), our confidence that there probably exists some reason for the pattern, so that it is not a coincidence, is well-grounded. It is extremely unlikely that it was not the case that the coin would probably continue this pattern in each instance (e.g., extremely unlikely that it was fair or weighted toward tails). For the same reason, whatever it may be, the coin will probably land heads the next time it is flipped.

Is the universe like a random number generator that through a colossal coincidence has given us vast patterns? Only if we believe this do we not have a reason to believe that inductive methods will continue to be successful. Mackie argued that no matter how confident you are that the universe is random, unless you are virtually certain, each day, when the apparent patterns continue (e.g. the Sun remains bright and the grass green), you should adjust your belief somewhat towards the opposite view that events at different times are probabilistically dependent (i.e. the universe is not random). Similarly, no matter how confident you are, a priori, arbitrarily or otherwise, that it is extremely unlikely on any given occasion that the Sun will rise, unless you are virtually certain, each new instance should push you further in the other direction. Such blind and unyielding confidence would not be attractive to a skeptic like Hume. All that is required is that we withhold judgment, rather than groundlessly assign probability 0 to the possibility of dependence. The skeptic should not dogmatically believe in the randomness of the universe any more
than in the existence of underlying causes.

While Laplace and others relied on the principle of indifference, Mackie showed that we do not need the principle of indifference to solve the problem of induction. We only need a modicum of doubt applied to the assumption that the Universe must be completely random, with no probabilistic dependence over time. As long as we are agnostic about whether the probabilities of distinct events are related, apparent patterns give us reason to believe such probability relations linking events at distinct times do exist – and we can help ourselves to an almost endless supply of these observations.

Inductive reasoning proceeds via coincidence-lessening explanation. Where patterns build up far past a point of statistical expectation based on any non-zero prior probabilities we assign, we should increase our expectation of probabilistic dependence in unobserved cases. “By how much should we increase our expectation of dependence?” may be an interesting question, but is not relevant for our purposes. Hume would have been satisfied with an argument that we should increase our expectation of probabilistic dependence in unobserved cases to any extent based on observation. Hume, from his radically skeptical standpoint, thought that an inductive premise is probabilistically irrelevant to (and epistemically independent of) inductive conclusions.

One might object that a solution to the problem of induction must “provide a guarantee that the probability of an inductive conclusion ever attains a degree at which it begins to be of use.” (Strawson 1963, 255) Positive relevance is not necessarily enough to establish this. It is unclear whether this additional requirement must be met
in order to fully solve the problem of induction. On the one hand, Hume’s discussion of inductive reasoning in infants suggests that he is interested in the justification, or lack thereof, of our actual, concrete inductive beliefs, which certainly involves something much stronger than judgments of mere positive probabilistic relevance. On the other hand, Hume believed on principled grounds, due to the separateness of the ideas of cause and effect, and the lack of either an *a priori* or non-circular experiential argument relating them, that an extrapolative inference could never have evidence in its favor, no matter how small. It seems that, at the very least, establishing positive probabilistic relevance on the basis of induction would chip away significantly at Hume’s skeptical foundation.

To summarize, if the powers we posit are not *timelessly* likely to persist, their persistence is no more likely than the persistence of the patterns they are invoked to explain would be in their absence. There would thus be no motivation for inferring that powers exist at all, if powers aren’t taken to be inherently likely to persist once they exist. Hume claims to be granting, for the sake of argument, that there were powers existing in the past; but the very reason for hypothesizing the existence of such powers implies that they must be inherently likely to persist, since otherwise they would explain nothing. This is why the Inference to the Best Explanation, Bayesian and statistical solutions to Hume’s problem of induction are not circular, as commonly charged - “explanations” that are not held to be mode-transcendent are explanatorily fruitless: they do not reduce the overall improbability of what we have to accept. In order to play its intended role, the dependence we appeal to must be timeless (more generally, tenseless), unlikely to stop at any particular time.79

---

79 Phrasing suggested by Thomas Ryckman.
4.2 Objection 1: Fair Sampling or Selection Bias?

P.F. Strawson poses an objection to the assumption of randomness, most clearly present in the sampling form of the solution. He writes that there is no “a priori guarantee that different mathematically possible samples are equally likely to be drawn.” (Strawson 1963, 254-255) Ironically, Foster also rejects the sampling version of the solution on the grounds that it assumes “that the sample was drawn at random,” meaning that “the method of selection gave equal chances of selection to all possible relevantly sized samples.” (Foster, 28) Foster says that this assumption is “crucial.”

In response, first, like Goodman’s new riddle, the concern that the sample might not be random is distinct from the problem of induction and is not, therefore, crucial (or even relevant). Hume claims simply that there is no justification for inferring unobserved data from observed data, writing, “we have no reason to draw any inference concerning any object beyond those of which we have had experience.” (Hume, T1.3.12, 139) He does not make an exception for cases in which the objects are chosen by controlled fair sampling. Hume believed, on principled grounds, that no extrapolative inference can be justified by reason. The absence of containment of the idea of the effect in the impression of the cause begins the search for an argument for induction from experience against which the circularity challenge is raised. Absence of containment of the unobserved in the observed, and possible differences between them, remain even under assumptions of randomness.

For example, looking at a world map, I observe 1000 points at random and discover that approximately 2/3 are blue, representing water. From this I can reasonably extrapolate that approximately 2/3 of the total points on the map are blue.
Although the assumption of randomness is stipulated/built in to the example, the inference remains extrapolative, from the observed to the unobserved. The observed and unobserved points remain wholly separate and distinct (from our perspective).

Similarly, placed in a house for a year on an alien planet, suppose I go outside on a random selection of days and observe that a nearby star can be seen every twenty-four hours. If my selection is large enough, I can reasonably infer that if I had stepped outside on another random day during the year, I probably would have seen the star on that day as well.\textsuperscript{80} In this case as in the former one, it seems that the assumption of randomness does not prevent the skeptical argument of the problem of induction from arising. Thus the randomness challenge involves a fundamentally different kind of skepticism than the problem of induction. Hume’s radical skepticism can be raised even in a random sampling scenario, so the objection to a supposed assumption of random sampling seems beside the point.

Second, the question is not simply whether the sample is biased or not, but whether, even if it is biased, we can conclude that the next case is likely to be biased in the same way. If so, then this bias is irrelevant to the question of the justification of singular inductive inference. When predicting the future based on past and present observations, any long-standing biases responsible for our past observations will probably continue to have the same effect on our observations of the future. We might worry that future observations will be picked from a different population than past observations. Strawson objects to the sampling solution, “No limit is fixed beforehand to... the multiplicity and variousness of different populations, each with different

\textsuperscript{80} Hume’s problem can be raised not just with respect to whether the Sun will rise tomorrow, but equally with respect to whether it rose on any random unobserved day in the past; induction applies to random retrodiction no less than to prediction.
constitutions, any one of which might replace the present one before we make the next draw.” (Strawson 2000, 234) But if the population were likely to be replaced at any given time, it is unlikely that the same population would have stayed in place for so long, until now.

Harrod makes this point: “Is this bias likely to come to a dead stop now?... It matters not whether the seeming regularity has been due to a real regularity or due to bias; either way... observables are likely to present themselves to us with their accustomed regularity.” (Harrod, 115) He concludes from this, correctly I think, that “the postulate that our sample of samples has been unbiased can be dispensed with.” (Harrod, 117) For example, if 90% of a large sample of voters polled so far voted for candidate A, this does not imply that it is likely that approximately 90% of the entire electorate voted for candidate A. Perhaps, for example, voters were chosen partly by some accidental feature correlated with voting patterns, like owning a landline telephone that pollsters are legally permitted to call, a form of selection bias against younger voters who often only use cell phones. Nevertheless, the probability that the next person who will be polled voted for candidate A is approximately 90%, because the same biased sampling method will probably continue to be used, given that it was used with such regularity in the past. Perhaps the sample is biased; but how is it that this exact bias could have persisted for so long? It must be that the bias has a propensity to persist, and hence is likely to persist into the next case.

If a coin is flipped many times and repeatedly lands heads, this is evidence in favor of the view that it is biased. But how can we know that it is likely to continue to be biased in the next case? If this bias is subject to change at all, there must have been
some second-order bias in favor of its continuation. If the bias in the coin could just as easily have ceased as continue at any time, the appeal to the bias is itself an appeal to an equally great coincidence. In order to avoid colossal coincidence, we must conclude that the coin is likely to continue to be biased for some length of time\(^{81}\), given the similarities between our past and present epistemic states. Similarly, there must be something keeping the law of gravitation going, even if gravitation itself is not time-impartially necessary. In order to avoid colossal coincidence, we must conclude that the persistence of gravitation is likely to continue for some length of time beyond any time it has long persisted, in the absence of cross-inductive information.

Third, the objection to the assumption of unbiased sampling is eliminated by the conjunction of Groundhog Day and the principle that epistemic probability supervenes on psychological structure. Bonjour writes:

> A skeptical hypothesis of the sort in question is just a special case of the general possibility... that the fact of observation might itself influence the evidence... such skeptical explanations are not alternatives to the straight inductive explanation of standard inductive evidence in the way that normal non-inductive hypotheses are. (Bonjour, 216)

Bonjour’s point is especially clear for Groundhog Day. Perhaps, one might argue, the observed is distinguished from the unobserved just in virtue of being observed; the sample is obviously biased in the sense that all observed instances share the common property of being observed. But, as Bonjour asserts, this is not the problem of induction. The character of being unobserved does not make the unobserved unlike the observed in any relevant sense, on Hume’s argument. We have a case of A, B, A,

---

\(^{81}\) In particular, the shorter the length of time, the more likely it is to continue for at least that long.
B, A, B, A, where we want to conclude B; not a case of A, B, A, B, A, B, A where we want to conclude B’ (i.e. B-in-the-unobserved). This is a different problem.

In the case of Groundhog Day, our present epistemic point of view is a random sample of the population consisting of our present and past epistemic points of view in virtue of being psychologically interchangeable with our past points of view. If all of our information now is qualitatively identical to all of our information at other times, then the same potential biases are equally likely in all of them. Psychologically interchangeable states may be reflections of distinct metaphysical realities with different objective probabilities, but these differences are fenced off from the epistemic states and therefore cannot alter the epistemic probabilities.

4.3 Objection 2: Bias Toward Marked or Unmarked States: Equally Likely?
I have argued that coincidences, i.e. marked unlikely occurrences, are unlikely in any particular case, making the disjunction of their alternatives more likely. This implies that any particular marked complex state is, a priori, more likely than any particular unmarked complex state, other things being equal.\textsuperscript{82} For example, of the three ten-letter series, CCCCCCCCCC is more likely than CCCCCCJCC, or XJRKDNGIWY. Perhaps the world is random, in which case they are all equally likely. But if the world is not random, bias towards any particular marked state is more likely than bias towards any particular unmarked state. What could justify this assumption?

I regard this not as an assumption, but as a consequence of the argument. The claim is that CCCCCCCCCC is more likely than CCCCCCJCC, because the latter is less marked. CCCCCCCCCC is part of the set consisting of BBBBBBBBBB and

\textsuperscript{82} This was brought to my attention by Mark Crimmins.
other marked series, while CCCCCCCJCC is part of the much larger set of less marked series. To assume that each unmarked state must be as likely as each marked state (or that any particular unmarked state must be as likely as the possibility of bias towards markedness) would be an improper use of the principle of the indifference, comparing apples and oranges. Nor should we use the principle of indifference to assign each possible degree of bias towards markedness equal prior probability, or anything so specific. Like Mackie, I instead claim that we should adopt a principle of tolerance with respect to the possibility of bias towards markedness.

The following consequence of the argument is roughly the same, though perhaps in a more intuitive form: \(a\ priori\), it is more likely that a process will last 1000 or more rounds than exactly 999 rounds, even if it is more likely that it will end eventually, and most likely that it will never begin in the first place. This is similar to Mackie’s claim that “The observation of a certain spread of uniformity raises the probability that the extent of uniformity is considerably greater than that spread much more than it raises the probability that that extent is equal to or only a little greater than that spread.” (Mackie 1979, 173) It is unlikely that we exist on the edge of a very long process. This is closely related to the Doomsday argument (we are likely to be relatively close to the middle of the number of humans who ever live - though this argument may not adequately take into account the entirety of our information), and to the claim I made in Chapter 3 that the laws of physics may change eventually, but will probably not change exactly now or at any other specific, randomly chosen moment. It is supported intuitively by the idea that to think that a process will last exactly 999 rounds, instead of 1000, 1001, 1002, or some other nearby larger number, is \textit{ad hoc}. 

135
Relatedly, focusing our attention on the previous rounds, the probability of continuity after 999 is roughly the same as the probability of continuity after 998, 997, and 996, given that the state of information is so similar. This probability must be great, or else it probably would not have continued for so long, in case after case.

4.4 Objection 3: Zero Prior Probability of Dependence?

In Chapter 3 and earlier in this chapter, I discussed J.L. Mackie’s alternative to the principle of indifference, namely a principle of tolerance, i.e. a non-zero prior probability of (probabilistic) dependence among events at different times. Mackie argues that any non-zero prior probability of dependence, no matter how small, is sufficient to launch his argument for time-transcendent dependence, given the observed regularities. But why should we accept this assumption of non-zero prior probability?

One response would be to argue that dependence is conceivable, and therefore possible, and therefore has non-zero probability. We can imagine different possible explanations of the interdependence of probabilities over time. For example, the apparent separateness of events at different times may, somehow, turn out to be illusory, such that when we observe an enduring object or repetitive event we are looking at a single entity from different “angles.”83 We have a concept of unity, so it seems possible that our perception of time as separating an object at different times is just a matter of our perception, and the difference between a rock at different times is like the difference between Cicero and Tully: a difference of perspective only, which could be seen from the point of view of God or an electron. Unity, which would

---

83 This suggestion is similar to what Mackie calls, “The thought of things as persisting through time and processes as projecting themselves in time.” (Mackie 1979, 175. Also quoted above, Ch 2.)
ground dependence, seems to be conceivable and therefore possible (though we cannot, perhaps, imagine exactly how it would hold, as we cannot see the whole picture).

However, Mark Crimmins points out that possibility does not entail non-zero probability. Considering a dartboard in a continuous space, “Since dead-center is just one point on the dartboard, and a single point has measure zero with respect to the area of the board, the chance that the dart will hit dead center is less than every positive real number.” (Mark Crimmins, email correspondence, 3/3/11) But still it is possible that it will hit dead center. Something can be possible yet still be one of an infinite number of possibilities, and thus be an insufficient foundation for an inductive argument based on Bayesian updating or the like. If the possibility of unity is like the possibility of hitting the center of the dartboard, the fact that unity is possible provides no assistance in our argument. But the possibility of unity or dependence over time is not analogous to possibility of hitting the center of the dartboard, because it is not a member of an infinite set of relevantly similar possibilities (as are the points on the dartboard, under the randomness assumption). Moreover, to assign each world exhibiting dependence equal probability to each world not exhibiting dependence would be an incorrect application of the principle of indifference, given that there are known differences in kind that may be relevant to the comparative probabilities.

Suppose that the argument for induction is successful unless interdependence has zero probability. Why should we worry that interdependence has zero probability? Given that interdependence is not impossible, there is no positive reason to think that it has zero probability. Compare the case where we find a coin on the street, flip it 100
times to find that it lands “heads” every time, and conclude that the coin is probably unfair - even if we have never seen or heard of an unfair coin. Of course, I need to begin with a non-zero probability that the coin is unfair to reach this conclusion; but why not? Why should I instead adopt a zero prior probability when it is clearly possible, given my information, that the coin is unfair? One should adopt non-dogmatic skepticism in the absence of rational persuasion, not groundless near-certainty. Before we have any evidence, we should “suspend judgment”, not stubbornly cling to confidence in non-unity. (Hume, E5.1) The fact that unity is possible does not strictly entail that unity has greater than 0 probability, but by removing impossibility, it removes one way in which the probability would be 0 (namely, if it were impossible), and leaves us without any prima facie reason for thinking that the probability is 0.

An assumption of non-zero probability is importantly different from an assumption that the future will probably be like the past, in the philosophical context of the problem of induction. It may help to see this clearly if we “invert” the probability range, focusing on the event that unity has zero probability instead of the event that it has non-zero probability. Adopting non-zero probability of unity is a matter of what Mackie calls “tolerance”, a seemingly reasonable epistemic starting point. But avoiding zero probability of unity or dependence is instead a matter of abstaining from stubbornness or groundless confidence/dogmatism, an attitude that the skeptic in particular ought to be sympathetic to.

Prior probability zero that the coin is unfair is the same as prior probability one that the coin is fair. Probability 1 is not, perhaps, the same as justified certainty, just as
probability 0 is not the same as epistemic impossibility. Crimmins writes that though the probability is 1 that the dart will not hit dead-center, “still one shouldn't be certain that the dart won't hit dead-center,” because it is (at least apparently) epistemically possible that it will. (Crimmins, email correspondence, 3/3/11) Thus it seems that justified certainty is stronger than probability 1. Still, we can turn the skeptical question around and ask, “What could justify us in being nearly certain that the coin is fair, assigning it probability 1?” Such confidence would be *ad hoc*. The problem of induction is not the same as the question: “What grounds do you have for not being extremely confident that the coin is fair, before it is ever flipped?”, or, analogously, “What grounds do you have for not being extremely confident that there is no probabilistic dependence in the world over time, before any observations are made?” But this is what the zero prior probability objection amounts to: a challenge to justify the initial choice of non-dogmatism, or tolerance towards the possibility of bias/dependence. Skeptical standards suggest not being blindly very confident in the absence of evidence or argument. Before experience, given no evidence or argument about whether the world contains trans-temporal dependence, we should not be blindly confident one way or the other.

Another objection, different from but related to the objection that the prior probability of dependence might be zero, is the objection that there may be no probability of interdependence (Compare: the distinction in geometry between “zero slope” and “no slope”). We only need to show that there is some probability *range* of dependence. But surely every possibility has a probability of some range between 0 and 1, inclusive. I argued in Chapter 3 that probability is limiting relative frequency,
though this does not constitute a “real definition” of probability (i.e. one that cuts to its “essence”). We might try to imagine a case in which a possibility has no limiting relative frequency, in order to make sense of the idea of a possibility with no probability. A limiting relative frequency is the hypothetical frequency that would be approached as the number of trials tends to infinity. Perhaps, we might imagine, after one thousand trials the frequency would be 1/2, after one million trials 8/9, after one billion trials 1/5, and so on, never stabilizing.

I do not know whether this idea of a possibility with no limiting relative frequency makes sense. Yet even in this example, the frequency always stays within the full range from 0 to 1, and so we can say that the event has a probability range, however broad. For nothing can happen more often than always or less often than never. Perhaps we cannot say anything more specific than that the limiting relative frequency (and thus the probability) must fall within this range, just as we can often only give a broad probability range in everyday life situations in which our information is not quantified. Such a probability range, however broad, if supplemented with a non-zero prior probability assumption, is equivalent to Mackie’s principle of tolerance and allows the argument for induction to get its footing.

A simpler response to both of these objections (i.e. zero prior probability and no prior probability) is that the supervenience of epistemic probability on psychological structure (within Groundhog Day) guarantees that the probabilities over time are not only interdependent, but, indeed, equal. This is an advantage of presenting the solution in epistemic terms. We do not need to worry that the probability of probabilistic dependence is 0 or non-existent, because it is known to be 1. We know
that the probabilities are equal, because the epistemic states they are relative to are intrinsically the same; *a fortiori*, we know that the probability that the probabilities are equal is neither 0 nor non-existent.
Conclusion: Induction and Causation

The method of inference to lesser coincidence has been used to argue for scientific realism, the existence of God, fine-tuning, other minds, an external world, and much else, including much everyday reasoning. Some of these arguments are intuitively more plausible than others. Though the arguments vary in degree of plausibility, they all seem to rely on the same reasoning that something must be the case in order for an observed pattern to constitute less of a coincidence. In the case of induction, the thing in question is a principle of uniformity (or perhaps something weaker, depending on how we define “principle of uniformity”).

Let us consider three additional instances of inference to lesser coincidence reasoning, which in my view rank in descending order of plausibility: the case for an objective direction of causation in time, for an unknown coincidence-lessening explanation of the low entropy state of the Big Bang, and for the related, absurd hypothesis that I am a Boltzmann brain.

First, the case for an objective direction of time seems strong. The idea is that the objective reality of the earlier-to-later direction of causation (in at least most of our observations) would explain the observed time-directionality of entropy and radiation, in the sense of rendering them less coincidental. Owens summarizes, “Causal explanation has a temporal direction because if we move through the causal nexus in the past-to-future direction, we find fewer coincidences than if we run through it in the opposite direction.” (Owens, 84) Salmon says the same: “Conjunctive forks that are open are always open to the future and never to the past. Since the statistical relations found in conjunctive forks are said to explain otherwise improbable coincidences, it
follows that such coincidences are explained only in terms of common causes, never common effects.” (Salmon, 99-100, in Pitt) Here, Owens and Salmon use the argument form of inference to the only alternative to colossal coincidence, used above in the justification of induction, to argue that earlier events causally explain later events, and not vice versa.

To help illustrate this point, consider Dummett’s apple, which falls from a tree branch, bounces a few times, rolls to a stop, and gradually disintegrates. Viewed in reverse, the forces of nature conspire to bring the earth together to form an apple and then cause it to roll up a hill, bounce, fly up to a tree, and attach to a branch. It appears to be a colossal coincidence, a conspiracy of seemingly independent occurrences, which leads to the unlikely creation of greater order (analogous to lower entropy). The earlier-to-later model of explanation contains fewer coincidences than the later-to-earlier model, and therefore it is more likely to be correct, vindicating commonsense. Despite the temporal symmetry of the physical laws, observation indicates that there is an asymmetry in the direction of causation in time. If the later-to-earlier description were objectively true, and not the earlier-to-later description, yet events occurred the same way, the world would be composed of many colossal coincidences, such as the regular occurrence of flying apples stopping exactly where a stem is waiting for them to attach. The world would be teleological, meaning that the correct direction of explanation, in terms of the objective world, would be reversed, despite appearances (and against all probability).

The arrows of entropy, radiation, etc., do not constitute, but rather indicate, the direction of causation in time. In essence, we think of the future as an epiphenomenon

84 Order is marked, so if unlikely, it is a coincidence.
of the past. “X is causally prior to Y where, if an intervention had been applied to X, Y would have been different, but not vice versa.” (Mackie 1979, 180) This is what is meant by the commonsense thought that time has a directional asymmetry. If, hypothetically, something were miraculously inserted into the space-time fabric at a certain point, we imagine that effects would propagate forwards in time but not backwards. This intuition is manifested in our ideas of time travel in science fiction, and miracles in religion. For example, when Marty McFly travels back in time from the 1980s to the 1950s in Back to the Future, his appearance in the 1950s does not have backwards propagating effects into the 1940s. But this can only be inferred from, not reduced to, the arrows of thermodynamics and radiation; it is logically consistent, though extremely unlikely, for causes to propagate towards the past in the sense outlined by this hypothetical intervention, even in a world in which the thermodynamic and radiation arrows are the same as in the actual world. Just as it is possible but unlikely for entropy to decrease into the future in an earlier-to-later world, it is possible but unlikely for entropy to increase into the future in a later-to-earlier world, and perhaps we live in such a world, though it is extremely unlikely. The observed arrows provide evidence in favor of our intuitive beliefs about the direction of causation in time, making probable the intuitive counterfactual that an intervention from outside of time would propagate only towards the future.

The laws of physics appear to be time-symmetric: as Loschmidt stated, if the velocity of every particle in the Universe is reversed, then according to the known laws of physics, these particles must trace their steps backwards to their original position at the Big Bang, seemingly in violation of the Second Law of Thermodynamics. However, this is only a consequence of time-symmetry, not a violation of it, as the laws of physics are still valid in the reverse direction. This is a fundamental aspect of the arrow of time, and it provides evidence for the direction of causation in time, making probable the intuitive counterfactual that an intervention from outside of time would propagate only towards the future.

Thermodynamics. Boltzmann showed that entropy must probably increase towards the future from a non-equilibrium state, presupposing a causal asymmetry in the direction of time. Without this presupposition, Boltzmann’s claim that entropy must probably increase toward the future would be subject to what Huw Price calls “the double standard fallacy.” (Price 1997) Without an objective asymmetry, we would be forced to conclude that entropy must probably increase towards the past as well as the future. Of course, it cannot increase at every point towards both the past and the future, so a paradox arises. The parity is broken and the paradox resolved if there is an objective direction of time, or a direction of causation in time, such that later events are correctly explained in terms of earlier events and not vice versa.

With regard to the low entropy of the Big Bang, which explains the present low entropy of the Universe, Huw Price argues that it too requires a coincidence-lessening explanation. This view strikes me as having some plausibility, although I think it is arguable whether the low entropy state of the Big Bang must be unlikely if unexplainable. Although there are many more ways for the entropy at the time of the Big Bang to have been higher than lower than it was, it is not clear that all of the possible initial states are equally likely in the absence of some further explanation; it is not clear that this would be a legitimate application of the principle of indifference.

Finally, if we accept that the low entropy of the Big Bang is extremely unlikely unless there is a coincidence-lessening explanation, and we also accept that there is no such explanation, then we can argue, based on inference to lesser coincidence reasoning, for the absurd conclusion that it is more likely that I am a Boltzmann brain arising from a random fluctuation within a high-entropy universe, than that I am the
embodied individual living in a low-entropy universe that I take myself to be. It seems to me that this is a sign of the absurdity of accepting both premises simultaneously (viz. that the low entropy of the Big Bang needs an explanation, and that it has none), not a problem for the inference to lesser coincidence form of reasoning. We should avoid the conclusion that the low entropy of the Big Bang is a colossal coincidence, whether or not this requires positing a further explanation. For this conclusion might indeed lead to the absurd consequence that I am more likely to be a Boltzmann brain than the person I take myself to be.

I have endorsed two arguments involving inference to lesser coincidence reasoning: for the probable success of induction, and for an objective asymmetry in the temporal direction of causal dependence. From these conclusions we can begin to give a description of the meaning of the concept ‘cause’. The cause/effect relation is whatever it is that makes (most of) our experiences of regularity in nature non-coincidental. For example, the fact that turning the key in the ignition causes the engine to start lessens the coincidentalness of the persistent conjunction of key-turnings and engine-startings. Moreover, ‘cause’ is distinct from ‘effect’ due to the different direction of explanation, which can be understood in terms of counterfactuals about interventions, as discussed above. Perhaps these properties of causation - having a direction, and lessening the coincidence of our observations of regularity - do not constitute the essence of causation. Our meaning is fixed by these properties, but may also depend on an extrinsic component, similar to the case of ‘water,’ which was secured as the referent ‘H2O’ before anyone was capable of knowing this to be the case. Just as we could not have known whether all water-like substance was in fact
water (H2O), as opposed to twater (XYZ), we do not now know whether ‘causation’ names all probabilistic dependence in the world. The dependence involved in quantum entanglement, for example, may turn out to be an exception. It is epistemically possible that there might be probabilistic dependence with a temporal arrow in the world other than that of causation, but this is an empirical question.
Bibliography


Salmon, Wesley, “Should We Attempt to Justify Induction?” Philosophical Studies. Springer Netherlands Volume 8, Number 3, April 1957.


Sober, Elliott, “Coincidences and How to Think About Them.” Online: philosophy.wisc.edu/sober/Coincidences%20July%201%202010.pdf


