What Are the Drivers of Induction? Towards a Material Theory+

Julian Reiss, Durham University

CHESS Working Paper No. 2019-05
[Produced as part of the Knowledge for Use (K4U) Research Project]
Durham University
July 2019

The K4U project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 667526 K4U) The above content reflects only the author's view and that the ERC is not responsible for any use that may be made of the information it contains

CHESS working paper (Online) ISSN 2053-2660

What Are the Drivers of Induction? Towards a Material Theory+

Julian Reiss

Julian Reiss
Department of Philosophy
Durham University
50 Old Elvet, Durham DH1 3HN
julian.reiss<at>durham.ac.uk

Induction is sometimes conceived as a method that leads, by means of mechanically applicable rules, from observed facts to corresponding general principles. […] Actually, however, no such general and mechanical induction procedure is available at present; otherwise, the much studied problem of the causation of cancer, for example, would hardly have remained unsolved to this day. Nor can the discovery of such a procedure ever be expected.
Hempel 1966: 14

1. Introduction

John Norton's Material Theory of Induction (Norton 2003, 2005, 2008, forthcoming) has a two-fold, negative and positive goal. The negative goal is to establish that formal logics of induction fail if they are understood as universally applicable schemas of induction. The positive goal is to establish that it is *material facts* that enable and justify inductive inferences. I argue in this paper that Norton is more successful with his negative than with his positive ambition. While I do not deny, and don't know anyone who does, that facts constitute an important type of enabler and justifier of inductions, they are by no means the only type. Below I will suggest that there are no less than six other types of background information scientists need and use to fuel and warrant inductions. The discussion of additional enablers and justifiers of inductions will show there are practically important and intellectually challenging methodological issues Norton's theory prevents us from seeing because it leaves out this or that type of enabler and justifier.

Norton provides three arguments in favour of the material theory of induction (Norton forthcoming: Ch.2, 1-2; emphasis original):

1.1. *Failure of universal schema*: […] no attempt to produce a universally applicable formal theory of induction has succeeded.
1.2. *Accommodation of standard inferences*: […] the successes of many exemplars of good inductive inferences can be explained by the material theory of induction.
1.3. *Inductive inference is powered by facts*: The **ampliative** character of inductive inference precludes universal schemas.

None of these arguments establishes the material theory on its own, and even jointly they are at best suggestive. That no formal theory of induction has succeeded so far does not mean that no formal theory ever will succeed. More importantly, taking the failure of past formal theories as evidence that no 'universally applicable formal theory' will succeed does not unequivocally speak in favour of Norton's material theory. It only speaks in favour of some theory that is not universally applicable and formal. That the material theory is able to explain the successes of many instances of good inductive inferences, again, provides evidence of its truth but leaves open the possibility that alternative theories explain instances of inductive success equally well or better. That inductive inference is 'powered by facts' allows the possibility of it not being powered by facts alone.

It is therefore possible to be largely in agreement with Norton's arguments and yet hold a different theory. The paper proceeds as follows. To pave the ground, I will describe the negative aspect of Norton's theory, largely approvingly, in the next section. In Section 3 will discuss his positive case for material facts as drivers of inductions. Section 4 will introduce a number of material elements Norton has left out in the Material Theory. Section 5 will add a number of *normative* enablers and justifiers of induction. Section 5 concludes.

Norton remains ambiguous between a normative and a descriptive reading of his theory. He certainly wants his theory to be descriptively accurate and maintains that in scientific practice, material facts play the role of enabling inductive inferences. But at times, he seems to be claiming more, viz. that material facts *in fact* warrant or justify inductions, quite independently of whether or not scientists are aware of this. I want to remain neutral on this issue and will, in what follows, refer to material facts as the *drivers* of inductions. Using this term I hope to convey a dual descriptive and normative meaning: that which does in fact enable inductive inferences and that which justifies or warrants it.

2. Formal Theories of Induction are Unsuccessful

Induction is a type of inference, the act of passing from one set of propositions, statements, or judgements taken to be true — the premisses — to another — the conclusion — the truth of which is believed to follow in one way or another from the truth of the premisses. Paradigm cases of good inferences are deductive inferences, which are such that the truth of the premisses *guarantees* the truth of the conclusion.

Some basic examples of deductive inferences include:

A. Modus Ponens

If it rains, then it pours.
<u>It rains.</u>
Therefore, it pours.

B. Disjunctive Syllogism

Either the gardener was the murderer, or the stable boy.
<u>The stable boy wasn't the murderer.</u>
Therefore, the gardener was the murderer.

C. Dilemma

If Antigone follows King Creon's order not to bury her brother, she'll betray her love for him and most deeply held values.
<u>If Antigone does secure a respectable burial for her brother, she'll be stoned to death.</u>
Therefore, Antigone will either betray the love for her brother and most deeply held values or be stoned to death.

Though I have used specific statements concerning concrete examples in each case, what makes deductive inferences special is that their validity derives fully from the logical *form* of the statement and is therefore independent of the statements' *content*. Modus Ponens, for instance, has the general form:

If A, then B.
<u>A</u>
Therefore, B

and the inference is valid no matter what we substitute for A and B, including obvious nonsense. Thus,

If pigs can fly, wallabies are larger than kangaroos.
<u>Pigs can fly.</u>
Therefore, wallabies are larger than kangaroos.

is a valid inference (even though, of course, from false premisses to a false conclusion in this case).

When Norton denies that all formal schemas of induction fail, he is in fact saying that induction does not work this way. In case of inductive inferences, the reliability of an inference is *not* invariant to substitutions of alternative premisses of the same form.[1] Let us take the simplest case of enumerative induction or inductive generalisation. That has the form:

> <u>Some As are B.</u>
> Therefore, all As are B.

This inference is reliable, for instance, if we substitute 'electron' for 'A' and 'has a mass of 9.10938356 × 10^{-31} kilograms' for 'are B', but not if we substitute 'gold coin' and 'has a mass of one ounce'. Sometimes weakening the conclusion helps to make an inference more reliable. Thus, 'Some ravens are black, therefore all ravens are black' is subject to exceptions due to the (rare) existence of albino ravens. Thus, weakening the conclusion to 'Almost all ravens are black' improves the inference. This won't work, however, when the As are cats rather than ravens.

The same point — that inductive inferences are not reliable merely in virtue of their form — can be made with respect to all theories or models of induction. An additional problem for simple enumerative induction is that it is very narrow in its applications. Plainly, many inductions are not inferences from 'some' to 'all' but rather inferences from effects to causes (e.g., when a disease is inferred from symptoms or a perpetrator from fingerprints and other clues) or from a collections of facts to a unifying hypothesis (e.g., when a hypothesis about dietary trends is inferred from facts about people's weights and other aspects of their health). Norton calls these cases of 'hypothetical induction' (e.g., Norton 2003, 2005).

The hypothetico-deductive account (e.g., Hempel 1966), for example, models inductive inferences as follows:

> Hypothesis H deductively entails evidence E.
> <u>E.</u>
> Therefore, H.

Formally, this is an instance of an inference called the 'fallacy of affirming the consequent'. The reason the inference is a fallacy in deductive logic is that the fact that H deductively entails E does not preclude the existence of one or more alternative hypotheses H', H'' etc. that equally entail the evidence E. Thus, many diseases produce similar symptoms, the same set of clues may have been left by many different suspects, many dietary behaviours may be responsible for the same patterns in health outcomes. To be reliable, hypothetico-deductivism must therefore be supplemented by mechanisms for selecting among the competing hypotheses.

Proponents of inference to the best explanation (IBE) maintain, unsurprisingly, that the hypothesis which best explains the evidence, is to be selected. One immediate problem for applying the model is that there are many ways to interpret 'explains' and 'best'. To make things simple, let us assume that 'explains' means 'causally explains' (and that we know what it means to explain causally) and that explanations in terms of causes that are more frequent or more likely to obtain are better than those that refer to rare or unusual causes.

Understood this way, IBE certainly works well for many cases. Even though both H = 'The patient has bronchitis' and H' = 'The patient has lung cancer' would causally explain a patient's episode of coughing, inferring H would be reasonable unless there are other symptoms that cannot be explained by H. But it doesn't work for other cases. Both Newton's gravitational theory and quantum mechanics violated widely held assumptions about the behaviour of causes prevalent at the time when they have come to be widely

[1] In logic, 'validity' is a technical term which refers to the truth-preserving property of an inference. Since inductive inferences are not truth preserving, it might confuse some readers to refer to an inductive inference as 'valid'. 'Reliable' is, however, not ideal either as it carries the connotation 'with high probability'. I do not mean to imply that a good inductive inference must confer a high probability on its conclusion. But there are no good alternative terms.

accepted (Day and Kincaid 1994). Newton's theory posits action-at-a-distance (in violation of the assumption that causes must be contiguous with their effects). Quantum mechanics is a genuinely stochastic theory (in violation of the assumption that causes are sufficient for their effects).[2] These cases show that theories that explain causally less well than alternatives were selected over these alternatives for reasons not directly related to explanation.

A third family of accounts of induction Norton distinguishes is the family of probabilistic accounts. Bayesian confirmation theory (e.g., Howson and Urbach 2006), degrees of belief are represented by probabilities, and the dynamics of belief change are governed by Bayes' Theorem. A scientific hypothesis H is assigned a probability P(H) = p. An observational statement E is evidence in favour of H if and only if the posterior probability of H conditional on E is higher than the prior, unconditional probability: $P(H \mid E) > p$. Bayes' Theorem states how to calculate the posterior probability:

$$P(H \mid E) = P(E \mid H)P(H)/P(E).$$

As Bayes' Theorem is indeed a theorem in probability theory, Bayesian inference is in fact deductive, and the force of a Bayesian argument stems entirely from the assignment of probabilities. This limits the applicability of Bayesian confirmation theory dramatically.

Decision theorists distinguish among situations of certain, of risk, and of uncertainty (e.g., Resnik 1987). Decisions under certainty are covered by deductive logic. Decisions under risk and uncertainty, which require inductive reasoning, are characterised by the existence of a well-defined outcome space and a probability measure over that space in the former and the absence of such in the latter case. The statistician Leonard Savage has referred to these types of situations as 'small worlds' and 'large worlds', respectively (Savage 1972).

In small worlds, when outcome spaces and probabilities are known, Bayesian confirmation theory works very well. Suppose a patient fears he might suffer from some disease and gets tested. H is the hypothesis that the patient has the disease, E the positive test result. The test has a known sensitivity, $P(E \mid H)$, of 95%, and a known specificity, $P(\neg E \mid \neg H)$, of 99%. The patient can be regarded as a randomly drawn individual from a population of patients in which 15% suffer from the disease. Using the expansion $P(E) = P(E \mid H)P(H) + P(E \mid \neg H)P(\neg H)$, we can rewrite Bayes' Theorem as:

$$P(H \mid E) = P(E \mid H)P(H)/[P(E \mid H)P(H) + P(E \mid \neg H)P(\neg H)],$$

and, filling in the numbers, calculate: 95%*15%/[95%*15% + 1%*85%] = 94.4%.

The question now is to what extent scientific inference can be modelled on medical testing or the analysis of games of chance. Bayesians scientific reasoning happens in large worlds but assume that they can be analysed, using simplifications and idealisations, as if they were small worlds. Norton and I are among the critics who argue that the assignment of probabilities can be highly misleading (and result in poor inferences) unless it is grounded in well-supported assumptions about the stochastic process responsible for the outcomes, as it is in games of chance and medical testing (e.g., Norton 2011, Reiss 2011, 2014).

Suppose a new bird has been discovered on some island and it happens to be black. How should we assign the prior probability to the hypothesis that all birds of this kind are black in the absence of any information about the bird's genus membership and ontogeny? Even using a principle of indifference (Keynes 1957 [1921]) that assigns equal probabilities to each possible outcome is inapplicable because absent other information, the outcome space is not clear. 50% (because there are two possible outcomes: black and non-black) is just as reasonable as 16.67% (because in opponent process theory there are six main colours) or any smaller number (as the light spectrum is continuous). Next we need to determine the likelihoods.

[2] A proponent of IBE might counter that neither contiguity nor sufficiency for the effect is a necessary component of the concept of cause. I will discuss the influence of conceptual considerations on induction below.

Presumably, if all members of this bird species are black, then we'd expect the one in front of us to be black and thus $P(E \mid H) = 1$. But what should we assume in any of the other cases for $P(E \mid \neg H)$? We'd need not only information about how many possibilities there are, also about how the individuals are distributed among the possibilities.

There are theorems showing that different assignments of numbers to priors eventually converge, but they do not solve the problem (cf. Norton forthcoming: Ch. 1). First, they are true only under unrealistic assumptions including the assumption that the incoming pieces of evidence are independent conditional on the falsity of the hypothesis. But of course, if evidence is collected repeatedly, it is collected under similar circumstances that are not independent from each other. Second, the priors wash out only in the long run but scientists tend not simply to repeat experiments or observations for the sole purpose of observing probabilities to converge.

3. Material Facts are Drivers of Induction

If formal approaches to inductive inference fail, it is a good idea to consider a material alternative. The terminology Norton uses is reminiscent of Rudolf Carnap's distinction between the formal and the material modes of speech, the former concerning the use of language, the latter, objects (or other entities) and their relations (Carnap 1937). But the formal/material mode distinction neither captures the difference Norton is after, nor does it make sense to say that objects play a role in inferences.

The formal/material distinction can also be found in the work of Wilfrid Sellars (Sellars 1953) and, inspired by Sellars, in Robert Brandom's work (Brandom 1994, 2000). Sellars and Brandom distinguish formal and material rules of inference. Formal rules of inference have been introduced above. Modus ponens et al. are valid in virtue of their form. Their validity is invariant to substitutions of the particular sentences that appear in the inference. Material inferences, by contrast, are valid in virtue of the meaning of the concepts that appear in the inference. For example, we can infer 'The streets are wet' directly from 'It is raining' without having to assume that the argument is an enthymeme, leaving out a premiss such as 'Whenever it is raining, the streets are wet', due to the content of the concepts involved.

Norton, however, explicitly rejects this approach, writing (Norton forthcoming: Ch. 2, 29):

> When I developed the material theory of induction, I was not aware of Sellars' and Brandom's notion of material inference and, in particular, Brandom's use of the term "material inference." […]
>
> The difficulty is that our notions of material inference differ slightly, as far as I can see. That means that it would have been better at the outset if I had chosen another name. For Brandom, the above inference is material since it is made good by the concepts invoked in the premises. In my view, it is material since I locate the warrant for the inference in the background material fact…

In this passage, Norton states that it is a 'background material fact' that warrants an induction. Unfortunately, he is not consistent in his description of the source of warrant. The original article that introduced the Material Theory (Norton 2003) speaks of 'material postulates' that 'license' or 'underwrite' inferences. But postulates and facts are two very different kind of thing. A postulate is an essential premiss in an argument. A postulate thus is a presupposition for an argument to go through but it may well be false. Social scientists standardly use the postulate of rationality in their analyses of social events but are aware of its idealising character.

Facts, by contrast, are 'that which is the case'. They are the truth-makers of propositions or the obtaining of states of affairs.[3] Facts are usually *contrasted*, not equated, with hypotheses.[4] A 'false fact' is either an

[3] There are many alternative philosophical theories of facts (Mulligan and Correia 2017). What I say about facts here is, I hope, neutral between these theories.

oxymoron or a category mistake. Social scientists do not use 'the fact of rationality' in their analyses of social events.[5]

Since the 2003 article defines: 'I shall call these licensing facts the material postulate of the induction' (Norton 2003: 650), some prominent counterexamples in the book notwithstanding, I will go with the facts reading and assume that the Material Theory maintains that facts rather than material postulates or substantive background assumptions are the drivers of inductive inferences.

The observation that all induction is local is the opposite side of the rejection of formal and universal theories of induction.[6] In their analysis of IBE, Day and Kincaid wrote (Day and Kincaid 1994: 282):

> IBE names an abstract pattern whose force and success depends on the specific background assumptions involved. Without substantive assumptions both about explanation in general and about specific empirical details, IBE is empty. In short, appeals to the best explanation are really implicit appeals to substantive empirical assumptions, not to some privileged form of inference. It is the substantive assumptions that do the real work.

The same can be said about inductive generalisation and Bayesian confirmation theory. We are happy to measure the mass of an electron only once or a few times (in case there are reasons to doubt the accuracy or precision of the measurement procedure) because we know, and we accept the Standard Model of particle physics as a background material fact, that elementary particles are homogenous in their intrinsic properties. We have no such fact to license the analogous inference in case of the gold coins. But this is entirely contingent on how the world is. If, for instance, there was a world-wide government monopoly on the minting of gold coins, the monopoly was strictly enforced and the government produced only sovereigns with a very reliable process, we could equally determine the mass of one or a small number of coins and infer immediately to all.

The point has already been made in the context of Bayesian confirmation theory. That theory works if and when the right kinds of facts about the stochastic process responsible for observable results are known. Thus, again, it is material facts, relevant to the case at hand, that drive inductive inferences.

4. Material Drivers Norton Left Out

Thus, I agree with Norton in his rejection of formal theories of inductive inference. However, what Norton puts in its place is wanting in several respects. While I agree that material facts are essential to inductive inferences, by assuming that ampliative inference is *exclusively* 'powered by facts' Norton's theory draws attention away from philosophical issues concerning inductive inference that are both practically important and intellectually challenging. These issues are: the role of theory in inductive inference, idealisation and adequacy-for-purpose, and the normative nature of inductive inference.

4.1. The Role of Theory in Inductive Inference

That it is facts that drive inductions makes the Material Theory untenable as a descriptive theory of induction. This is because scientists frequently have to rely on theory or 'postulates' the truth of which is at

---

[4] Facts are contrasted with theories or hypotheses on the one hand and values on the other. This section focuses on the former contrast. I will address the influence of values on inference below.

[5] The reason why I belabour this point will become apparent in the next section.

[6] This view is in no way novel or specific to Norton's theory. In a symposium on scientific inference, Ron Giere commented on the formal nature of the Bayesian approach (Giere 1997: S183):

[S]trictly speaking, there is no logic underlying scientific inference. There are only methods with various desirable operating characteristics. These operating characteristics, being sensitive to the experimental context, lack the universality of logical principles. But that is just what makes them well-suited to the job of acquiring reliable experimental knowledge.

least disputed if not openly denied in their inferential practices. Norton's book is full of examples. In Chapter 1, for instance, René-Just Haüy's crystallographic theory, which is false as it assumes each crystalline substance to have a single characteristic crystallographic form, plays such a role. Norton notes (Ch. 1: 19):

> This is the crudest version of how chemists pass from a single sample to all. What is notable is that it is no inductive inference at all. The inference is deductive and authorized by early crystallographic theory.
>
> Of course this is an extreme case and a purely deductive passage was possible only during a brief window of a few decades of the early years of Haüy's crystallographic theory. The theory soon encountered anomalies.

Here, then, we have a theory rather than a fact that powered an inference.[7] While this is very often the case, everywhere in science, it comes particularly to the fore at the boundaries of sciences, when novel phenomena are encountered and investigated — as in early crystallography.

Economics has witnessed a debate about the role of theory in inference that is almost as old as the discipline itself. This is the debate between groups of economists I have called 'Ricardians', who maintain that inference should always proceed against a theoretical model that is needed to select, order, and interpret evidence, and their opponents, whom I have called 'Baconians', who reject theory as unreliable and therefore urge to generalise from the facts very gradually and without the benefit of theory (Reiss unpublished). The first instalment of this debate was between the actual Ricardians (a group of classical economists) and the Cambridge Inductivists (a group around William Whewell); the current instalment is between economists following a structural or Cowles Commission approach and those following a design based approach.

In the 1940s the debate circled around the role of theory in business cycle research. In a famous review of Arthur Burns' and Wesley Clair Mitchell's *Measuring Business Cycles* (Burns and Mitchell 1946), Dutch-American economist Tjalling Koopmans distinguished between a 'Kepler stage' and a 'Newton stage' of inquiry, the former aiming to discover 'empirical regularities', the latter at 'fundamental laws'. The laws of the Newton stage are more fundamental because they are at once more elementary and more general (Koopmans 1947: 161). While Koopmans acknowledged Burns and Mitchell's contribution to the Kepler stage of inquiry in the field of economics, he maintained (*ibid*.: 162; emphasis original):

> that in research in economic dynamics the Kepler stage and the Newton stage of inquiry need to be more intimately combined and to be pursued simultaneously. Fuller utilization of the concepts and hypotheses of economic theory (in a sense described below) *as a part of the processes of observation and measurement* promises to be a shorter road, perhaps even the only possible road, to the understanding of cyclical fluctuations.

To support his judgement, Koopmans provided three arguments:

My *first argument*, then, is that even for the purpose of systematic and large scale observation of such a many-sided phenomenon, theoretical preconceptions about its nature cannot be dispensed with, and the authors do so only to the detriment of the analysis. (*ibid*.: 163; emphasis original)

This, then, is my *second argument* against the empiricist position: Without resort to theory, in the sense indicated, conclusions relevant to the guidance of economic policies cannot be drawn. (*ibid*.: 167; emphasis original)

---

[7] Norton might of course maintain that Haüy's theory was in fact used, but that crystallographers were not justified in using it for inductive inferences, i.e., that the inferences were not warranted. However, he makes no suggestion to the effect that crystallographers at the time merely felt, but weren't in fact, justified in making the inference, or perhaps that anyone who shares their background beliefs would be justified in making the inference relative to that constellation of beliefs.

[A]ny rigorous testing of hypotheses according to modern methods of statistical inference requires a specification of the form of the joint probability distribution of the variables. […T]he the extraction of [useful] information from the data requires that, in addition to the hypotheses subject to test, certain basic economic hypotheses are formulated as distributional assumptions, which often are not themselves subject to statistical testing from the same data. Of course, the validity of information so obtained is logically conditional upon the validity of the statistically unverifiable aspects of these basic hypotheses. The greater wealth, definiteness, rigor, and relevance to specific questions of such conditional information, as compared with any information extractable without hypotheses of the kind indicated, provides the *third argument* against the purely empirical approach. (*ibid*.: 170; emphasis original).

I will talk about statistical inference in detail below.

The reason to revisit this debate here is that researchers in domains that aren't already settled, especially when its phenomena are complex and the capacity for experimentation is limited, face exactly the dilemma that characterised the Burns/Mitchell-Koopmans exchange. Without theory, data cannot be selected, ordered, interpreted or indeed used for inductive inferences. But since there is no widely accepted theory, it is regarded by critics as unfit for the job. Use theory? Damned if you do, damned if you don't.

This is not the place for a resolution of the dilemma faced by researchers at the frontiers of science (for an attempt, see Reiss unpublished). Let me make just two remarks. First, the dilemma is a genuine one that is not resolved trivially. In many sciences data are exceedingly easy to come by but exceedingly hard to use as a basis for effective inferences. Theory would solve many inferential problems but there is no theory that is universally or even widely accepted. Second, Norton sides, without argument, with the radical inductivists or 'Baconians' in the debate, those who wanted to learn gradually from experience alone. The problem is that, at least in economics, the purely inductivist approach has never been executed with much success. Now, this may well be due to accidents of history.[8] But the history of science indicates that other disciplines too have profited from background assumptions or 'postulates' that go well beyond the known facts and have thus helped to turbocharge inductions that would not have advanced much in their absence. The interesting question for a methodologist who wants to contribute to a resolution of debates among practitioners such as the above is the question to what extent and in what manner a background postulate can violate the facts without losing its ability to power inferences effectively. By limiting the drivers of induction to facts, Norton loses the ability to address this issue.

Thus: (1) *theories* are additional drivers of induction.

4.2. Idealisations and Adequacy-for-Purpose

A related but by no means identical issue is the widespread use of idealisations in scientific inferences. Theories are bodies of substantive hypotheses used to systematise and unify a range of diverse phenomena. Idealisations are more specific hypotheses that conflict with known facts (or are presumed or suspected to do so) but that are useful nevertheless. Theories may well contain idealisations, but the two are not the same.

We have already encountered an example of an idealisation above: the routine use of the assumption of rationality in social research. There is no doubt that in many contexts, social scientists are justified in using the assumption. Milton Friedman, for instance, argued that if businessmen did not behave as if they maximised profits, they'd be driven out of business (Friedman 1953). The assumption can thus be used to model the behaviour of business leaders unless specific good reasons to think otherwise can be given (e.g., because short-run behaviour is being analysed, incentive structures aren't appropriate, there is significant market failure etc.).

[8] As I have argued in Reiss 2008.

What is clear is that there is no fact of rationality that could be used to warrant inferences. Of course, it is not just social scientists who idealise. Norton, for instance, discusses the cosmological principle, according to which the spatial distribution of matter in the universe is homogeneous and isotropic when viewed on a large enough scale, as providing warrant for inferences. The cosmological principle is probably an idealisation but most certainly an assumption rather than a known fact.

The use of idealising assumptions in inference comes into sharp relief in *statistical* inference. Statistical inferences always proceed against a *probability model* (e.g., Hoover 2003). Probability models are representations of the data-generating process from which the analysed data set was sampled and contain assumptions about the functional form of the relationships among the sampled variables, the distribution of the error term (which measures the net influence of omitted variables) as well as the sampling mechanism. Frequently made assumptions include random sampling, linearity, normality and that errors are independent and identically distributed (IID).

Many such assumptions are plainly false. Many samples are convenience rather than random samples, for instance. To build a model that 'works', i.e., that is simple enough so that existing statistical tools can be brought to bear on the problem at hand and yet realistic enough so that inferences are not too far off the mark is an art. This is in part due to the fact that relatively small differences between probability model and data-generating process can lead to significantly different inferences. Consider the following example due to David Freedman (Freedman 2009: 28):

> Suppose, for example, that in a certain jurisdiction there are 1084 probationers under federal supervision: 369 are black. Over a six-month period, 119 probationers are cited for technical violations: 54 are black. This is disparate impact, as one sees by computing the percents: In the total pool of probationers, 34% are black; however, among those cited, 45% are black.
>
> A t-test for "statistical significance" would probably follow. The standard error on the 45% is $\sqrt{.45 \times .55/119}$ = .046, or 4.6%. So, t = (.45 − .34)/.046 = 2.41, and the one-sided P is .01. (A more sophisticated analyst might use the hypergeometric distribution, but that would not change the outlines of the problem.) The null hypothesis is rejected, and there are at least two competing explanations: Either blacks are more prone to violate probation, or supervisors are racist. It is up to the probation office to demonstrate the former; the t-test shifts the burden of argument.
>
> However (ibid.: 29-30): Suppose the citation process violates the independence assumption in the following manner. Probation officers make contact with probationers on a regular basis. If contact leads to a citation, the probability of a subsequent citation goes up, because the law enforcement perspective is reinforced. If contact does not lead to a citation, the probability of a subsequent citation goes down (the law enforcement perspective is not reinforced). This does not seem to be an unreasonable model; indeed, it may be far more reasonable than independence.
>
> More specifically, suppose the citation process is a "stationary Markov chain." If contact leads to a citation, the chance that the next case will be cited is .50. On the other hand, if contact does not lead to a citation, the chance of a citation on the next contact is only .10. To get started, we assume the chance of a citation on the first contact is .30; the starting probability makes little difference for this demonstration.
>
> Suppose an investigator has a sample of 100 cases, and observes seventeen citations. The probability of citation would be estimated as 17/100 = .17, with a standard error of $\sqrt{.17 \times .83/100}$ = .038. Implicitly, this calculation assumes independence. However, Markov chains do not obey the independence assumption. The right standard error, computed by simulation, turns out to be .058. This is about 50% larger than the standard error computed by the usual formula. As a result, the conventional t-statistic is about 50% too large. For example, a researcher who might ordinarily use a critical value of 2.0 for statistical significance at the .05 level should really be using a critical value of about 3.0.

Thus, a difference that is significant under the assumption of random sampling turns out not to be significant under a Markov chain model. What is important to note is that neither the independence nor the Markov chain assumption represents a material fact of the citation process. The Markov chain model might be more realistic but it remains an idealisation the adequacy of which has to be assessed in the light of the purpose of the inference. What is good enough for one purpose may be hopelessly inadequate for another.

Statisticians sometimes point out that assumptions such as IID can be tested (e.g., Spanos 2010). This is true, of course, but it doesn't help with the present problem. Such tests are statistical tests and thus equally proceed against probability models that contain large numbers of significant idealisations. As long as we use modern statistical tools in inductive inference, we won't get around the problem of idealisation.

As above, there is a methodological issue lurking here that is as important to practitioners as it is challenging to philosophers of science: how do we determine, especially in the absence of knowledge of the 'true model', whether the falsehood we are using is good enough for the purpose at hand?[9] Sometimes we will be able to determine which idealisations have worked with hindsight, but are there any ways to tell before the fact which idealisations are likely to work? Again this is an issue the Material Theory prevents us from seeing clearly because of its exclusive focus on facts as drivers of inductions.

Thus: (2) *idealisations* and (3) *purposes* are additional drivers of induction.

5. The Normative Nature of Inductive Inference

Purposes are, of course, already normative elements in a more complete Material Theory+ of induction. In this section I will add more normative drivers. Specifically, ethical norms, methodological norms, and conceptual norms, will be shown to play significant roles in inductive inferences. We will again also see how Norton's exclusive focus on the material facts of induction prevents us from seeing important methodological issues.

5.1. Ethical Norms

One major argument about how ethical norms enter inductive inference is quite old and very well known among philosophers of science. I would also say that it is widely accepted among (contemporary) philosophers of science, but Norton explicitly rejects it. So let us rehearse the argument and address Norton's criticism.

The argument is, of course, the argument from inductive risk that was introduced in Richard Rudner's 'The Scientist Qua Scientist Makes Value Judgments' (Rudner 1953). The argument, in a nutshell, is the following. Inductive inference always involves a risk of error.[10] The error is of two possible types. A scientist can accept a hypothesis that is in fact false (a 'false positive'), or he can reject a hypothesis that is in fact true (a 'false negative'). There is a trade-off relationship between the two types of error, as one can be controlled completely at the expense of the other. If one never accepts new hypotheses, the risk to accept a false hypothesis is zero but one is certain to miss out many true hypotheses and vice versa. Scientists therefore have to make up their minds how best to trade off the two types of risk. Rudner now argues that the decision should be made on a consideration of the relative severity of the consequences to which each type of error is likely to lead. What is worse: poisoning or killing patients with drugs that aren't safe or foregoing the benefits of new treatments that are? Risking a planet-destroying chain reaction or foregoing the benefits of having a weapon with which fascism could most certainly be snuffed out? Finally, it is value judgements that guide scientists' assessments of the importance of the consequences.

Norton rejects the argument on two grounds (Norton forthcoming: Ch. 5, §5). First, he argues that these kinds of value judgements are rarely made in scientific practice. Most research is too far away from

[9] Parker 2009 makes some advance on this issue.

[10] Strictly speaking, I would argue that ampliative inference always involves what I'd like to call *inductive uncertainty* rather than risk. As we have seen above, the difference between uncertainty and risk is that in the latter case, outcome spaces and probabilities are known. Apart from well-designed and executed randomised trials, few methods generate probabilities and so situations of inductive risk are in fact very rare.

potential applications so that considerations concerning consequences are moot. Norton argues, second, that Rudner equivocates between two senses of the word 'scientist'. In the narrower sense, according to Norton, 'a scientist is merely someone who investigates nature, reporting what bearing the evidence has, with indifference to the broader human ramifications' (Norton forthcoming: Ch. 5, 10). Even when a hypothesis (e.g., about the safety of a drug or the absence of a planet-destroying chain reaction) has potential consequences, there mere acceptance or rejection of it does not. A scientist in the broader sense is 'someone who practices science and monitors the import of his or her work within the wider human society' (*ibid*.: 11). When acting as a scientist in this broader sense, her actions have of course important consequences and thus should be guided by value judgements. However, according to Norton, virtually all the work of scientists proceeds in the narrow mode.

I have a certain sympathy for Norton's first point. It is plainly not the case that scientists (individuals or groups[11]) are always in the position to anticipate the consequences of accepting hypotheses. He is of course right to say that the acceptance of the hypothesis that electrons are half-spin particles involves evidence and not value judgements. But I disagree that 'Virtually all the work of scientists proceeds in this mode' (*ibid*.). Virtually all of social research has direct implications. And it's not just social science. The same is true of much of psychology, biomedical research and epidemiology, engineering, AI and computer science, environmental science and climate science. There are two factors that influence whether values play a role in inference. One is human interest. The more we are interested in a research result going one way or another, the more likely will value judgements play a role in inductive inferences to the results (Dupré 2007). That the hypothesis that electrons are half-spin particles can be accepted without thinking too much about values has little to do with the scientific nature of the hypothesis and everything with the fact that the result does not matter to us in any way.[12] The second factor is effect size. Even though the problem of inductive uncertainty obtains in every case, when effects are huge, making an error of either type is so highly unlikely that the influence of value judgements is minimal. Effects in the sciences mentioned above tend to be quite small, however, and so the issue of how to trade off the two types of error remains an important one.

Whether hypotheses the truth or falsehood of which matters to us and small effect sizes are frequent or rare is an empirical question that cannot be determined by philosophical analysis. It seems to me that this kind of research is not too infrequent. But the point is: since it exists, a theory of induction that is able to accommodate values is more generally applicable than one that is not able to do so.

I am happy also to accept Norton's distinction between the narrow and the broad sense of a scientist but disagree that in the vast number of her actions, a scientist can be absolved from taking responsibility. Even if at the end of the day it is regulators and policy makers who translate scientific findings into regulations and policies, their actions essentially rely on scientific advice. A scientist contributing to a consensus, say, about anthropogenic climate change or the safety and efficacy of a new drug is as responsible for the consequences of a regulation or policy, to the extent that these consequences are foreseeable, as the regulator or policy maker because she co-determines the decision.

In a recent paper I have argued that normative considerations are among the 'pragmatic criteria' used to infer a hypothesis from the evidence. Specifically (Reiss 2015: 356; emphasis original):

> *Economic and other normative considerations: take into account economic and other costs and benefits when deciding to stop or continue probing the indirect support for a hypothesis*. Causal inquiry does not come for

[11] Harry Collins and Robert Evans argue in a recent book that the argument from inductive risk fails in part because individual scientists do not accept hypotheses but produce research results (Collins and Evans 2017). They advocate the establishment of a group of experts called 'The Owls' tasked with reviewing all the evidence concerning some topic and coming to an assessment of the consensus on that topic. The Owls would use value judgements in their assessment, thus the scientists don't have to.

[12] To be more precise, it does matter to us that we know the spin number of electrons but it does not matter what that number is.

> free. There are direct, opportunity, and ethical costs. These costs have to be traded off against the benefits of reducing uncertainty. The benefits of reducing uncertainty consist in the reduced chance of accepting a false or rejecting a true hypothesis. There are no strict rules on how to optimize the trade-off, and people holding different values will differ in their assessments. What is clear, however, is that a reasonable trade-off will seldom entail an indefinite continuation of challenging the indirect support for a hypothesis.

At a higher level of resolution, the a 'default-and-challenge' rule plays an important role. Many scientific communities adopt community-wide standards for trading off the two types of error, often the injunction not to accept more than 5% false positives. Individual scientists accept it as the default rule.[13] But if there are case-specific reasons to believe that the standard will lead to poor results in the given case, it should be amended. For example, if a new disease appears that is particularly deadly, it might be reasonable to loosen the standard temporarily, as it might make sense to tighten it up for drugs that do not promise much medical benefit (so-called 'me-too drugs').[14]

Thus: (4) *ethical norms* are additional drivers of inductions.

5.2. Methodological Norms

As pointed out above, all statistical inference proceeds against probability models. While this is true of all modern statistical inference, different statistical paradigms dictate different rules of inference, require the making of different sets of assumptions and have different endpoints. Consider the debate between classical and Bayesian statisticians about stopping rules. Suppose a scientists offers a statistician a set of 100 IID (and normally distributed) observations and asks her to test the hypothesis that the population mean is different from zero (see Berger and Wolpert 1988: 74 for this example). The sample mean is .2. Is this evidence against the null hypothesis?

Classical and Bayesian statistics give different instructions for how to proceed in addressing the question. A classical statistician will have to ask why the scientist stopped after collecting 100 observations. Depending on the answer, she will draw different inferences. The result might be significantly different from zero under one stopping rule but not under another. This is because different stopping rules define different outcome spaces, and in classical statistics the full outcome space enters the calculation of the test statistic. By contrast, the Bayesian test statistic depends only on the likelihood ratio. Stopping rules are therefore irrelevant.

The issue of stopping rules is controversial and more complex than suggested by this simple example (see for instance Mayo 1996, Mayo and Kruse 2001, Steel 2003, Steele 2013). But what the example shows is that different statistical paradigms license different inferences, *holding the background of material facts fixed*. The material facts of this case are not disputed between classical and Bayesian statisticians. And yet, classical and Bayesian statisticians will (generally) use different sets of inputs and different inference rules and come to different conclusions.

Importantly, the paradigm constrains the kinds of questions that can be addressed legitimately with its resources. Another bone of contention between classical and Bayesian statisticians is the base-rate fallacy. Bayesians have accused classical statisticians of committing the fallacy, i.e., of ignoring the relative sizes of population subgroups when assessing the likelihood of contingent events involving these subgroups (Howson 1997, Howson and Urbach 2006). Classical statisticians respond that the example that appears to show that classical testing involves an instance of the fallacy in fact has none of the features of a classical test (Spanos 2010). What is uncontroversial is that classical tests license inferences only about the properties of the populations from which the data were sampled; Bayesians make inferences about the probabilities of hypotheses.

---

[13] So far I agree with Levi 1960.

[14] 'Me-too drugs' have little to do with the 'me-too movement' but are similarly controversial. For more on these drugs, see Reiss 2010, Reiss and Kitcher 2009.

There is no material fact in the world that could help us determine whether classical or Bayesian statisticians are right about any of these matters. Arguments in support of either (or any other) paradigm involve normative considerations about the appropriateness of methodological standards as well as the desirability of goals and purposes of the inquiry (Steel 2005). Without such normative input, inductions could not get off the ground, at least not in modern statistics.

Thus: (5) *methodological norms* are additional drivers of inductions.

5.3. Conceptual Norms

Conceptual norms can play a very similar role as the methodological norms discussed in the previous subsection. They influence the informational requirements for an inference and what can be inferred. But material facts alone do not determine the appropriateness of conceptual norms.

The concept of cause is a case in point. Consider the following remarks made by Jacob Henle, a nineteenth century German physician, about causes in medicine (Henle 1844: 25, quoted from Carter 2003: 24):

> Only in medicine are there causes that have hundreds of consequences or that can, on arbitrary occasions, remain entirely without effect. Only in medicine can the same effect flow from the most varied possible sources. [...] This is just as scientific as if a physicist were to teach that bodies fall because boards or beams are removed, because ropes or cables break, or because of openings, and so forth.

Henle wrote in defence of the germ theory of disease according to which causes were necessary universal conditions for their effects (i.e., the diseases in question). Thinking about causes in this way was extraordinarily successful in the second half of the nineteenth century and has led to the discovery — and eventually treatment — of many diseases. But towards the end of the century the theory ran into anomalies, essentially due to cases where a cause appeared to be present but not the disease and vice versa.

Conceptual norms help to determine what kind of evidence is relevant to the evaluation of a hypothesis. If a cause is a necessary universal condition for its effect, then a given factor can be ruled out as a cause for a given effect if there are cases in which the effect is present and the cause is not and vice versa. It also tells us what kinds of inferences are licensed. Again if a cause is a necessary universal condition, we would expect, for instance, the effect to disappear after the cause has been eliminated.

Material facts determine whether a factor of interest (such as a microorganism) is a cause given a concept of cause but they do not determine which of a number of alternative concepts to accept in the first place. This is because a cause is, first and foremost, a useful factor. Michael Scriven, for instance, argues (Scriven 1966: 256):

> When we are looking for causes, we are looking for explanations in terms of a few factors or a single factor; and what counts as an explanation is whatever fills in the gap in the inquirer's or reader's understanding.

I would use a broader set of purposes but agree with the general point: a cause is any factor that is useful in view of certain kinds of purposes such as explanation, prediction, intervention, diagnosis of failure, attribution of praise and blame. Material facts of course help to assess whether a given factor can be used, say, to predict or explain outcomes. But it is also norms concerning the desirability of these goals and purposes and what their attainment means in a given context that shape our standards of conceptual adequacy.

Thus: (6) *conceptual norms* are additional drivers of inductions.

All three examples of norms as drivers of inductions discussed in this section lead to what is called 'factvalue entanglement' in science (Putnam 2002, Reiss 2017). Again, there are exciting methodological questions to be asked in this context but that will be ignored when the focus is on material facts as exclusive drivers of inductions: Is it a good idea to reduce the influence of values to a minimum (for instance, by ignoring normative drivers of inductions)? If values are a necessary element in inductive inference (or scientific practice more generally), how do we decide which sets values to use? How do we manage the influence of values in science? Which constituencies should scientists respond to when deliberating about the proper role for values in inference?

6. Conclusions

What I hope to have established in this paper are the following three claims:

A. Norton is correct in his negative claim that formal theories of inductive inference fail.
B. Norton is also correct in his positive claim that material facts play the role of drivers of inductions.
C. But Norton is incorrect in assuming that material facts are the *only* drivers of inductions. There are at least six additional drivers, *viz*.
    i. Theories
    ii. Idealisations
    iii. Purposes
    iv. Ethical Norms
    v. Methodological Norms
    vi. Conceptual Norms.

Any viable Material Theory+ of induction will have to incorporate these elements.

Bibliography

Berger, James and Robert Wolpert 1988. *The Likelihood Principle*. 2nd ed. Haywood (CA), Institute of Mathemathical Statistics.

Brandom, Robert 1994. Making It Explicit: Reasoning, Representing & Discursive Commitment. Cambridge (MA), Harvard University Press.

Brandom, Robert 2000. Articulating Reasons: An Introduction to Inferentialism. Cambridge (MA), Harvard University Press.

Burns, Arthur and Wesley Clair Mitchell 1946. *Measuring Business Cycles*. New York (NY), National Bureau of Economic Research.

Carnap, Rudolf 1937. *Logical Syntax of Language*. London, Keagan, Paul, Trench, Trubner & Co.

Carter, K. Codell 2003. The Rise of Causal Concepts of Disease: Case Histories. Aldershot, Ashgate.

Collins, Harry and Robert Evans 2017. *Why Democracies Need Science*. Cambridge, Polity Press.

Day, Timothy and Harold Kincaid 1994. "Putting Inference to the Best Explanation in Its Place." *Synthese*(271-95).

Dupré, John 2007. Fact and Value. *Value-Free Science? Ideals and Illusions*. Harold Kincaid, John Dupré and Alison Wylie, Eds. Oxford, Oxford University Press**:** 27-41.

Freedman, David 2009. Statistical Models and Causal Inference: A Dialogue with the Social Sciences. Cambridge, Cambridge University Press.

Friedman, Milton 1953. The Methodology of Positive Economics. *Essays in Positive Economics*. Chicago, University of Chicago Press.

Giere, Ron 1997. "Scientific Inference: Two Points of View." *Philosophy of Science* **64**(Supplement. Proceedings of the 1996 Biennial Meetings of the Philosophy of Science Association. Part II: Symposia Papers): S180-184.

Hempel, Carl 1966. *The Philosophy of Natural Science*. Upper Saddle River (NJ), Prentice-Hall.

Henle, Jacob 1844. "Medicinische Wissenschaft Und Empirie." *Zeitschrift für rationelle Medizin* **1**: 1-35.

Hoover, Kevin 2003. "Nonstationary Time-Series, Cointegration, and the Principle of the Common Cause." *British Journal for the Philosophy of Science* **54**: 527-551.

Howson, Colin 1997. "Error Probabilities in Error." *Philosophy of Science* **64**(Supplement. Proceedings of the 1996 Biennial Meetings of the Philosophy of Science Association. Part II: Symposia Papers): S185-194.

Howson, Colin and Peter Urbach 2006. Scientific Reasoning: The Bayesian Approach. 3rd ed. Peru (IL), Carus.

Keynes, John Maynard 1957 [1921]. *A Treatise on Probability*. London, Macmillan.

Koopmans, Tjalling 1947. "Measurement without Theory." *Review of Economic Statistics* **29**(3): 161-171.

Levi, Isaac 1960. "Must the Scientist Make Value Judgments." *Journal of Philosophy* **57**: 345-357.

Mayo, Deborah 1996. Error and the Growth of Experimental Knowledge. Chicago, University of Chicago Press.

Mayo, Deborah and Michael Kruse 2001. Principles of Inference and Their Consequences. *Foundations of Bayesianism*. David Corfield and Jon Williamson, Eds. Dordrecht, Kluwer**:** 381-404.

Mulligan, Kevin and Fabrice Correia 2017. Facts. *Stanford Encyclopedia of Philosophy*. Edward Zalta, Ed. URL = <https://plato.stanford.edu/archives/win2017/entries/facts/>.

Norton, John 2003. "A Material Theory of Induction." *Philosophy of Science* **70**(4): 647-670.

Norton, John 2005. A Little Survey of Induction. *Scientific Evidence: Philosophical Theories and Applications*. Peter Achinstein, Ed. Baltimore (MD), The Johns Hopkins University Press**:** 9-34.

Norton, John 2008. Must Evidence Underdetermine Theory? *The Challenge of the Social and the Pressure of Practice*. Martin Carrier, Don Howard and Janet Kourany, Eds. Pittsburgh (PA), Pittsburgh University Press**:** 17-44.

Norton, John 2011. Challenges to Bayesian Confirmation Theory. *Philosophy of Statistics*. Prasanta Bandyopadhyay and Malcolm Forster, Eds. Dordrecht, Elsevier.

Norton, John forthcoming. The Material Theory of Induction.

Parker, Wendy 2009. "Confirmation and Adequacy-for-Purpose in Climate Modelling." *Aristotelian Society Supplementary Volume* **83**(1): 233-249.

Putnam, Hilary 2002. The Collapse of the Fact/Value Dichotomy and Other Essays. Cambridge (MA),

Harvard University Press.
Reiss, Julian 2008. Error in Economics: Towards a More Evidence-Based Methodology. London, Routledge.
Reiss, Julian 2010. "In Favour of a Millian Proposal to Reform Biomedical Research." *Synthese* **177**(3): 427-447.
Reiss, Julian 2011. Empirical Evidence: Its Nature and Sources. *Handbook of Philosophy of Social Science*. Ian Jarvie and Jesús Zamora-Bonilla, Eds. Thousand Oaks (CA), SAGE**:** 551-576.
Reiss, Julian 2014. "What's Wrong with Our Theories of Evidence?" *Theoria* **29**(2): 283-306.
Reiss, Julian 2015. "A Pragmatist Theory of Evidence." *Philosophy of Science* **82**(3): 341-362.
Reiss, Julian 2017. "Fact-Value Entanglement in Positive Economics." *Journal of Economic Methodology* **24**(2): 134-149.
Reiss, Julian unpublished. The Perennial Methodenstreit: Observation, First Principles, and the Ricardian Vice. Durham University,
Reiss, Julian and Philip Kitcher 2009. "Biomedical Research, Neglected Diseases, and Well-Ordered Science." *Theoria* **24**(3): 263-282.
Resnik, Michael D. 1987. *Choices: An Introduction to Decision Theory*. Minneapolis, University of Minnesota Press.
Rudner, Richard 1953. "The Scientist Qua Scientist Makes Value Judgments." *Philosophy of Science* **20**(1): 1-6.
Savage, Leonard J. 1972. *The Foundations of Statistics*, Courier Dover Publications.
Scriven, Michael 1966. Causes, Connections and Conditions in History. *Philosophical Analysis and History*. William Dray, Ed. New York (NY), Harper and Row**:** 238-264.
Sellars, Wilfrid 1953. "Inference and Meaning." *Mind* **62**(247): 313-338.
Spanos, Aris 2010. "Is Frequentist Testing Vulnerable to the Base-Rate Fallacy?" *Philosophy of Science* **77**: 565-583.
Steel, Daniel 2003. "A Bayesian Way to Make Stopping Rules Matter." *Erkenntnis* **58**: 213-227.
Steel, Daniel 2005. "The Facts of the Matter: A Discussion of Norton's Material Theory of Induction." *Philosophy of Science* **72**: 188-197.
Steele, Katie 2013. "Persistent Experimenters, Stopping Rules, and Statistical Inference." *Erkenntnis* **78**: 937-961.