Causal Inference as Inference to the Best Explanation

Barry Ward

University of Arkansas

Abstract

We argue that a modified version of Mill's method of agreement can strongly confirm causal generalizations. This mode of causal inference implicates the explanatory virtues of mechanism, analogy, consilience, and simplicity, and we identify it as a species of Inference to the Best Explanation (IBE). Since rational causal inference provides normative guidance, IBE is not a heuristic for Bayesian rationality. We give it an objective Bayesian formalization, one that has no need of principles of indifference and yields responses to the Voltaire objection, van Fraassen's Bad Lot objection, and John Norton's recent objection to IBE.

1. Introduction

Inference to the Best Explanation (IBE) is commonly held to crucially involve explanatory virtues such as consilience, simplicity, mechanism, and analogy. Accounts of causal inference, on the other hand, generally do not. Mill's methods, for instance, make no appeal to such virtues. So, surprisingly, it seems that we shouldn't look to causal inference for guidance in constructing an account of IBE. Indeed, Steven Rappaport (1996) has criticized Peter Lipton's (1991, 2004) well known account on precisely the grounds that it's an account of causal inference *as opposed to* a theory of IBE.

This should give us pause. When we infer a causal generalization from our data, we commonly infer a generalization that explains aspects of that data: if we observe that a variety of chemical kinds of sodium salt are uniformly accompanied by a yellow flame when ignited, and infer that all sodium salts burn yellow / cause a yellow flame when ignited, we infer a causal generalization that explains the observed uniformity. Moreover, if we make that inference using something like Mill's methods, we use a methodology that is specific to causal / explanatory generalizations. If anything might be expected to provide an exemplar of IBE, such an inference should.

In this paper we'll show how causal inference can implicate the explanatory virtues and be reasonably construed as IBE. In particular, we'll show how something much like Mill's method of agreement has a place for the virtues and give the resultant methodology a Bayesian formalization.

Proponents of IBE commonly advocate that explanatory considerations merely provide a heuristic for realizing the judgements of an ideally rational Bayesian agent: IBE is descriptive of our actual practices, but only Bayesian considerations are normative. Since causal inference provides an independent source of normativity, our theory is not a heuristic one. Further, its causal basis eliminates any need for principles of indifference, distinguishing it from existing non-heuristic views. Thus, we obtain a novel, objective Bayesian confirmation theory. We briefly evaluate its relation to several influential confirmation theories, and argue that it deals with the

Voltaire objection, van Fraassen's Bad Lot objection, and a recent objection to IBE due to John Norton.

2. Causal Inference, Mechanism, and Analogy: Refining Mill's Method of Agreement

Mill (1882, 280) states his method of agreement as:

> "If two or more instances of the phenomenon under investigation have only one circumstance in common, the circumstance in which alone all the instances agree, is the cause (or effect) of the given phenomenon."

We can immediately discern a role for mechanism in this method. When we apply it to infer a cause, we infer a mechanism or part of one. Further, since any set of instances will have an arbitrarily large number of circumstances in common, to apply it we must restrict ourselves to some set of potentially causally relevant circumstances or factors i.e., at least partially specified potential mechanisms.

Mill intended his method to cover causal generalizations that are not restricted to the observed cases, for example (1882, 278) "the combination of an oil and an alkali causes the production of a soap." We can discern a role for analogy in such applications, as we shall now illustrate for a simpler example.

Suppose we're interested in whether sodium fluoride uniformly burns yellow / causes a yellow flame when ignited. However, it occurs in several different crystal forms and we initially believe that such variations might be causally relevant to flame color. We observe the ignition of the various crystal forms and find they invariably yield a yellow flame. The effect of interest is the color of the flame, the crystal forms are potentially causally relevant dissimilarities between cases of sodium fluoride, and, for reasons we'll shortly clarify, we'll take the method not as identifying the cause, but rather, restricting it to some single cause that satisfies the disjunction *being sodium fluoride or something invariantly associated with that*. We may reason as follows:

(1) The observed kinds of crystal form of sodium fluoride uniformly burn the same color

(2) So, the observed kinds of crystal form are not causally relevant dissimilarities.

(3) Those dissimilarities exhaust the potential causally relevant dissimilarities between cases of sodium fluoride (observed and unobserved).

(4) So, all sodium fluoride is relevantly causally similar.

(5) *Being sodium fluoride or something invariantly associated with that* specifies the relevant causal similarity between cases of sodium fluoride if there are no relevant dissimilarities.

(6) So, *being sodium fluoride or something invariantly associated with that* causes color of burning for sodium fluoride.[1]

We not only observe that our sodium fluoride samples burn the same color, but (7) that they burn yellow. Thus, from (6) and (7), we infer (8) and also (9):

(7) All observed sodium fluoride burns yellow.

[1] The claim is not that the disjunction is the cause, but that something that satisfies a disjunct is the cause.

(8) So, all unobserved sodium fluoride burns yellow

(9) So, *being sodium fluoride or something invariantly associated with that* causes yellow burning.

The argument from (1) to (2), and hence, from (2) and (3) to (4), establishes that all realizations of sodium fluoride are analogous/similar in regard of factors causally relevant to flame color. It's not an argument *from* analogy, one on which we argue from some similarity/analogy to a further similarity. It's an argument from the data *to* analogy, so to speak. The argument from (4) and (5) to (6) partially identifies the respect in which they are analogous, and hence, delimits the set of unobserved cases covered by the subsequent deductive argument from analogy between the observed and unobserved cases, the argument from (6) and (7) to (8), and the closely related deductive argument from analogy from (6) and (7) to (9).

Why infer that *being sodium fluoride or something invariantly associated with that* specifies the cause, as opposed to, for instance, *being sodium fluoride*? The fact that each crystal structure is observed to burn yellow only justifies inferring that those variations are not causally relevant dissimilarities between instances of sodium fluoride, not that the cause of the yellow burning is *being sodium fluoride* as opposed to something more general, such as *being a sodium salt*. That would minimally require observations of the burning of some sodium salts that are not sodium fluoride.[2] Of course, background beliefs about potential causal similarities may select

[2] Similarly, observed agreement won't justify inferring that the cause is *being sodium fluoride* as opposed to some aspect of the electronic structure of sodium fluoride, for instance, or indeed the electronic structure of sodium salts more generally.

one element of the disjunction as the cause. However, recognizing the disjunction as the only conclusion warranted by such data clarifies our understanding of its justificatory limitations.

When we say "invariantly" we mean invariantly across the range of cases implicitly specified by premise (3) i.e., the range of cases for which we hold the observed potential dissimilarities exhaust the potential dissimilarities. Depending on our background beliefs, that may be invariance across all nomologically possible cases—in which case we confirm a causal generalization that strongly resembles a law—or some more restricted set of possibilities, in which case we confirm a more parochial causal generalization. So, the scope of "all" in (4), (7), and (8), should be dictated by the context for (3).

The above chain of inference uses the same kind of data and—in tandem with relevant background beliefs—yields the conclusions licensed by the method of agreement.[3] However, it's more demanding in one respect. Mill's method doesn't include any requirement corresponding to premise (3), and so, licenses inferring that all sodium fluoride burns yellow even if we only observe a modest subset of the known kinds of crystal structures of sodium fluoride and reasonably retain our belief that each of the unobserved kinds of crystal structure might relevantly causally differ from those observed. (3) is required if we're to assign a high probability to—hereafter, *well confirm*—the causal generalization.[4] Let's now specify how we might reasonably justify (3).

[3] Also, its relative complexity makes it unsurprising that Mill treated it as one, *sui generis* kind of inference. Notably, the data that the various observed crystal forms burn yellow is used in two different ways. To justify (2), we only exploit the information that they burn the same color. However, in (7) we use the fact that they burn yellow for the argument from analogy.

[4] Mill (1882, 279) viewed his method of agreement as inferior to his method of difference, providing only modest support for its conclusions. With the addition of (3) and our disjunctive specification of the cause, we think more highly of it.

What kinds of cases must our evidence cover? There should be a kind for each factor that we believe might be a causally relevant dissimilarity. In this case, each crystal structure specifies one such kind. Also, if we allowed that other, *unknown* factors in addition to crystal structure might well be causally relevant to flame color, we could not be confident that, as per (3), the observed dissimilarities exhaust the potential causally relevant dissimilarities. Hence, the relevant kinds must be ones we're highly confident subsume only cases that are relevantly causally similar—this feature makes such kinds the natural "ground floor" for applications of the method of agreement. So, the kinds of cases that our data must cover to well confirm "Being a C or something invariantly associated with that cause Es" are causally individuated kinds of Potential Disconfirmers (hereafter, "kinds of PDs") i.e., kinds for which we have:

(i) A high degree of confidence that the subsumed cases are relevantly causally similar,

(ii) A single individuating factor that we believe might be causally relevant to the production of disconfirming cases / cases where E is not realized.

So, to justify (3) we must justify holding that our set of kinds of PDs is exhaustive and also our confidence that the cases subsumed by each kind are indeed relevantly causally similar.

How do we do this i.e., justify holding that there isn't some unknown factor or factors that cause not-E cases? Mechanism can provide crucial assistance here. Granted suitable background beliefs we might reasonably hold that the set of variations in crystal structure exhausts the potentially causally relevant variations in sodium fluoride. If, in addition, we have the theoretical resources to determine which crystal structures of sodium fluoride are possible, we can rationally judge that we have an exhaustive set. Even if we're not that sophisticated, we can still be in a position to make that judgement. If we've a good idea of the variations in conditions that can result in differing crystal structures—e.g., concentration and temperature of the solution from which crystals form—and on broadly varying those conditions have only found some finite set of variations in sodium fluoride crystal structure, we can be reasonably confident those variations are exhaustive. Thus, our specification of potentially relevant mechanism, in this case crystal structure, allied to knowledge of the ways in which variations in such mechanism can be caused, can delimit the set of possible causally relevant variations, providing good reason to believe we have an exhaustive set of kinds of PDs. So, mechanism has this additional, important role.

The kind of argument we used to confirm the sodium fluoride causal generalization is much more powerful than we might suspect. We can use another version to well confirm a causal generalization that covers an unlimited set of kinds of PDs i.e., an unlimited set of kinds of cases we initially hold might be relevantly causally dissimilar. Let's shift our focus to sodium salts more generally, and now assume we believe that only variations in chemical kind might be relevant to flame color—variations in crystal structure etc. are, by hypothesis, believed irrelevant.

In terms of modern chemistry a sodium salt is formed by the union of a sodium ion with the anion formed from an acid by removing a hydrogen ion. One family of sodium salts consists of those formed by the union of a sodium ion with a carboxylic acid with one hydrogen removed. The simplest member of this family has the molecular structure HCOONa (i.e., a hydrogen bonded to a complex consisting of a carbon and two oxygens with one oxygen bonded to a sodium ion). The next, H-CH_2-COONa, is similar, but with an extra unit consisting of a carbon bonded to two hydrogens inserted into the chain. The others differ only in the number of CH_2 units inserted: H-CH_2-CH_2-COONa, H-CH_2-CH_2-CH_2-COONa, and so on. There's no limit to the number of CH_2 units that can be inserted to generate novel chemical kinds that belong to this family of sodium salts, and so it's not feasible to observe members of each chemical kind. Nevertheless, we can justify well confirming that all sodium salts in the family burn yellow, with an argument that parallels our earlier one:

(1) For a suitable variety of values of n, Molecules of form $H(CH_2)_nCOONa$ uniformly burn the same color.

(2) So, the differences in molecular structure generated by arbitrarily varying the value of n are not causally relevant dissimilarities.

(3) Such dissimilarities exhaust the potential relevant causal dissimilarities between sodium salts of form $H(CH_2)_nCOONa$ (observed and unobserved).

(4) So, all sodium salts of form $H(CH_2)_nCOONa$ are relevantly causally similar.

(5) *Being a molecule of form $H(CH_2)_nCOONa$ or something invariantly associated with that* specifies the relevant causal similarity between cases of $H(CH_2)_nCOONa$ if there are no relevant causal dissimilarities among them.

(6) So, *being a molecule of form $H(CH_2)_nCOONa$ or something invariantly associated with that* causes color of burning for such sodium salts.

(7) All observed sodium salts of form $H(CH_2)_nCOONa$ burns yellow.

(8) So, all unobserved sodium salts of form $H(CH_2)_nCOONa$ burn yellow.

(9) So, *being of form $H(CH_2)_nCOONa$ or something invariantly associated with that* causes yellow burning.

What's distinctive here is the justification of (2) by (1). We can't observe cases of each kind. However, as we gather data regarding increasingly complex members of this family and find they invariably burn the same color, at some point we can very reasonably cease to believe the chemical differences among that family of sodium salts might be causally relevant dissimilarities, and indeed become highly confident they're not. So, the argument goes through as before, but now justifies a conclusion regarding an unlimited family of sodium salts.

As this example shows, posited mechanism can expand our inductive reach in a spectacular way: our justification of (2) by appeal to (1) explicitly depends on our specification of the molecular structure that covers this unlimited family of sodium salts: it allows us to evidentially cover the range of potentially causally relevant variation in the unobserved cases (arbitrary variations in number of CH_2 units) with the range in the observed cases (significant variety in number of CH_2 units).[5]

This kind of inductive argument is valuable in its own right. It doesn't have to be the only game in town to be an important part of our repertoire of rational inferences. However, it's unclear that corresponding inductions are justifiable without appeal to such arguments. A random sampling argument is only rational if we can justify holding our sample representative. Since no finite sample will include every kind of these sodium salts, justifying representativeness demands we justify holding that the unobserved kinds do not *relevantly causally differ* from the observed ones. It's unclear how we might do that without the use of mechanism that gets us

[5] We've now identified two ways in which mechanism can facilitate our modified method of agreement: ensuring that we have an exhaustive set of kinds of PDs, and facilitating inductions like this one. For a general discussion of the uses of mechanism in induction, see Thagard (2021).

from (1) to (2), and hence to (4), justifying representativeness.[6] An enumerative induction with any pretense to rationality similarly demands a justification of representativeness. So, the mechanism-based expansion in our inductive reach is not obviously independently available.

How does our account relate to other conceptions of analogy? Paul Thagard (1978) provides the most developed account of IBE in terms of the explanatory virtues.[7] He (1978, 89-91) focuses on analogy between existing well understood processes and novel explanatory hypotheses. By exhibiting the analogy between artificial and natural selection, Darwin (1962, chapter 1) rendered the latter explanations more comprehensible, as did Huygens (1950) his wave theory of light by exhibiting its analogy to antecedently understood wave aspects of sound. Such analogies also provide some evidential support for the novel theory. Similarities between the modifications of species that occur in selective breeding and in the wild, and the fact that the former is explained by (artificial) selection, gives some reason to take (natural) selection seriously as an explanation of the latter.

These arguments from analogy are more tentative than ours. As Darwin and Huygens explicitly recognized, they're prosecuted in the face of potential, unknown relevant dissimilarities—they didn't know enough about the mechanisms of natural selection and light to be highly confident there were none. These arguments, and Thagard's account of them, neither include a premise that there are no relevant dissimilarities, nor anything corresponding to the arguments *to* analogy that precede our arguments from analogy and justify our holding that there are no relevant dissimilarities between the cases. We should also note that Thagard's arguments

[6] For a related discussion of random sampling, see Ward (2022).
[7] Thagard has written extensively on IBE in the intervening years. However, this paper still neatly encapsulates his view of the virtues.

infer similar causes from similar effects; our arguments from analogy—those that go from (6) and (7) to (8) and (9)—infer similar effects from similar causes.

There are two other important differences. As illustrated, Thagard's are primarily inter-theoretic arguments. Further, while he (1978, 91) does not see explanation as reduction to the familiar, he recognizes that when we invoke this species of analogy, it's resemblance to the familiar that counts. Our arguments to and from analogy are intra-theoretic, and far from being a conservative force, the posit of novel mechanism such as molecular structure can, as illustrated, facilitate inductive confirmation that extends over hitherto uncharted realms.

Thagard's inter-theoretic arguments from analogy manifestly have an important scientific role. However, our intra-theoretic arguments have a different one. As we shall now see, the difference is reflected in our account of consilience and simplicity.

3. Causal Inference: Consilience and Simplicity

William Whewell (1847, 65) originated the term "consilience":

> "… the evidence in favor of our induction is of a much higher and more forcible character when it enables us to explain and determine cases of a *kind different* from those which we contemplated in the formation of our hypothesis…[I] will term it the *Consilience of Inductions*."

Thagard (1978, 79) offers two alternative explications:

"…let FT_i be the set of classes of facts explained by theory T_i. …(1) T_1 is more consilient than T_2 if and only if the cardinality of FT_1 is greater than the cardinality of FT_2, or (2) T_1 is more consilient than T_2 if and only if FT_2 is a proper subset of FT_1."

Neither Whewell nor Thagard provides an analysis of *kind of cases* / *classes of facts*. Thagard (1978, 80) observes:

"The most difficult feature…is the notion of a class of facts… the problem is merely pragmatic, concerning the way in which, in particular historical contexts, the scientific corpus is organized. The inductive logician must take this organization as given..."

The methodology we're developing from the method of agreement, hereafter, the Relevant Causal Similarity Methodology (RCSM), can explicate this notion and also explain why consilience has confirmational value, something neither Whewell nor Thagard attempted.

The confirmation of "*being of form $H(CH_2)_nCOONa$ or something invariantly associated with that* causes yellow burning" provides an exemplary case. Classes of facts are kinds of PDs: kinds of cases that, as per condition (ii) in our definition, we believe are *potentially* relevantly causally / explanatorily dissimilar from each other, but, as per condition (i), we're confident are individually relevantly causally / explanatorily similar. In this case, they're the chemical kinds of such sodium salts specified by differing values of n. A consilient generalization provides a single explanation of why all the subsumed chemical kinds burn yellow. By contrast, a non-consilient generalization provides distinct explanations of the yellow burning of the subsumed chemical kinds: the conjunction of the causal generalizations that each explain the yellow burning of one subsumed chemical kind in terms of it being of that particular kind or something specifically associated with it, shares the same set of kinds of PDs, but is non-consilient.

As shown above, the consilient generalization can be well confirmed by our successive arguments to and from analogy. However, the non-consilient one cannot, since it denies the analogy.[8] So, this notion of consilience has clear confirmational significance.

Indeed, in cases where there's a sizeable number of kinds of PDs, even excellent evidence that cases of each kind conform to the non-consilient generalization will, for Preface paradox like reasons, likely not well confirm it. Further, in our example, where there's an unlimited set of kinds of PDs, such evidence is unobtainable. So, consilience can be of the first importance in confirmation: non-consilient hypotheses may not even be significantly confirmable by data that would well confirm a consilient competitor. On the RCSM construal, the analogical source of consilience's confirmational significance is evident, and its importance for strongly confirming highly general causal hypotheses can hardly be overstated.

We can generally characterize consilience as follows: a causal generalization is consilient if, and only if, it provides a single explanation that covers its set of kinds of PDs. We shall also adapt Thagard's two criteria to make comparisons of degree of consilience *among consilient hypotheses*. The above hypothesis that unifies only the sodium salts formed from carboxylic acids is less consilient than "being a sodium salt or something invariantly associated with that causes

[8] The original method of agreement is similarly inapplicable. Agreement across the observed cases only licenses inferring the consilient generalization that contradicts the non-consilient one.

yellow burning", since its set of kinds of PDs is a proper subset of the set for the latter.[9] We might also use comparisons of cardinality in cases where one set of kinds of PDs is not a subset of the other. Both notions may provide a rough comparison of the confirmational gains that can be achieved in virtue of the compared generalizations' consilience.

Our account of consilience importantly differs from Thagard's. For us, consilience is explanatory unification and the degree of consilience of a consilient hypothesis is a measure of the scope of that unification. Thagard, however, only has a notion of degree of consilience, and that merely tracks explanatory scope without regard for unification. Thus, in the above example, the consilient and non-consilient generalizations cover the same classes of facts, and hence, by either of his definitions are equally consilient. However, mere scope is cheap—as per this example, we can get it by conjunction—and plausibly doesn't yield the dramatic confirmation that motivated Whewell's idea of consilience. Irrespective of what Whewell intended, our notion of consilience has clear confirmational significance. On that alone, it merits a place among the explanatory confirmational virtues.

Finally, we turn to simplicity. For Whewell (1847, 68), it's determined by the character of the hypotheses invoked to accommodate the data:

"…*simplicity*…the new suppositions resolve themselves into the old ones, or at least require only some easy modification of the hypotheses first assumed…In false theories…The new suppositions

[9] Incidentally, we might plausibly well confirm the latter by a sequence of such analogical arguments. First, we well confirm that each of the various chemical families of sodium salts burn yellow; then, we form an argument by analogy across the set of such families. So, the confirmational gains of consilience can be iterated.

are something altogether additional;--not suggested by the original scheme; perhaps difficult to reconcile with it."

In a similar vein, Thagard characterizes it as inversely related to the size of the set of auxiliary assumptions, A, required by T to explain the set of facts F, where an auxiliary (1978, 87) "is a statement, not part of the original theory, which is assumed in order to help explain one element of F, or a small fraction of the elements of F." He introduces a comparative notion on which we adjudicate between Theories T_1 and T_2 by comparing the associated sets of auxiliary hypotheses AT_1 and AT_2. However, as he recognizes, it's unclear how to make that comparison. We can't just count sentences, since we can trivially amalgamate all the auxiliaries in each set into one conjunction, and we can't use a subset relation, since competing theories will commonly employ entirely distinct sets of auxiliaries.

The RCSM provides an account that fits Whewell's and Thagard's reasonable concern with auxiliaries but doesn't involve counting them. Consider again the burning of sodium salts with molecular structure $H(CH_2)_nCOONa$. If we find for numerous values of n that these sodium salts uniformly burn yellow, we can reasonably cease to believe that variations in the value of n provide reason to suspect disconfirmation of "all such sodium salts burn yellow", and can well confirm it. Suppose instead, we find a mysterious anomaly: salts for which n = 17, 34, 51, and 68 burn orange. To explain such data we introduce an auxiliary, "for all n = 17. m, with m a positive integer, $H(CH_2)_nCOONa$ burns orange". We're also compelled to replace our initial hypothesis by "for all $n \neq 17.m$, all salts of form $H(CH_2)_nCOONa$ burn yellow". If we find that for increasing values of n, we still get uniform yellow burning for $n \neq 17.m$, and uniform orange burning otherwise, we might still reasonably well confirm each of our two causal generalizations by analogy. However, since we don't have a single subsuming explanation for the two kinds of cases, we can't well confirm their conjunction in that way. So, at best, our explanatory theory for this family of sodium salts consists of the conjunction of two well confirmed hypotheses, and other things being equal, that gets a lower probability than the single well confirmed hypothesis that was our initial hope. More generally, the more an explanatory hypothesis or theory fragments the domain to be explained, the less strongly our data can confirm it, ceteris paribus.

The arbitrariness problem has gone away. Simplicity is not about number of auxiliaries, but about the number of distinct explanations invoked to explain a phenomenon across a given set of kinds of cases. Thus, it's inversely related to the number of domain-specific laws or causal generalizations invoked. Hence, consilience and simplicity are closely, and inversely, related: a consilient explanation of a given set of kinds of cases will be simpler than a non-consilient one.

4. Objective Bayesian Formalisation and Principles of Indifference

The probability of a hypothesis, h, conditional on evidence, e, and background beliefs, K, is given by Bayes's theorem:

$$P(h|e \,\&\, K) = P(e|h \,\&\, K) \,.\, P(h|K) \,/\, P(e|K)$$

According to Bayesians, on learning e, a rational agent updates their personal probability in h from their prior, $P(h|K)$, to their posterior, $P(h|e \& K)$. Thus, the values on the right dictate their new probability for h on learning e.

Proponents of IBE have typically sought some kind of marriage with Bayesianism. The most popular view—variously advocated by Lipton (2004), Okasha (2000), McGrew (2003), Bird (2017), Dellsén (2018), and others—is that it's a heuristic used to (defeasibly) approximate rational Bayesian inference: IBE may often describe our actual inferential practices, but all normativity flows from Bayesian considerations. Since rational causal inference provides a distinct source of normativity, we reject this view. Our account requires an objective Bayesian formalization with the constraints on acceptable probability distributions dictated by causal / explanatory considerations. We'll now articulate these constraints.

4.1 Confirming a Causal Generalization holds for a kind of PD

We begin with the justification of premise (1) in our revamped Millian arguments. We justify (1) by first well confirming, for each kind of PD, that cases of that kind burn yellow. By definition, we're highly confident that cases of a kind of PD are relevantly causally similar. Hence, a modest number of suitable cases should well confirm conformity to an associated causal generalization:

(RCS) Let background beliefs, K, determine that K_i is a kind of case for which we have a high degree of confidence that the cases are causally similar in regard of factors relevant to property B. Let C_i be the causal generalization "being an entity of kind K_i or something invariantly associated with that causes the entity to be B", and P(.) be a probability distribution with background beliefs K. Then, if $e_1, e_2, ..., e_n$ consists of confirming data of the corresponding kind, i.e., each is a statement of the form "K_i a & Ba", then P(C_i| e_1 & e_2 &...& e_n) must be high / well confirmed, for even modest values of n.

This is an inductive rationality constraint on our priors: barring unusual background beliefs or confidences, such data should not undermine our high confidence in the relevant causal similarity of cases of that kind. Hence, our confidence in the specified subsuming causal generalization conditional on such data should be high, and conditionalizing on the data well confirms that generalization. So, observing a modest number of uniformly yellow burning sodium fluoride samples of crystal structure X well confirms that being sodium fluoride of crystal structure X or something invariantly associated with that causes a yellow flame, and observing a modest number of samples of, say, $H(CH_2)_4COONa$ well confirms that being $H(CH_2)_4COONa$ or something invariantly associated with that causes a yellow flame.

(RCS) doesn't directly impose a constraint on our priors for causal generalizations i.e., their probability conditional on our current background knowledge, K, alone. However, it presupposes we're dealing with a kind of case for which we have a high degree of confidence in relevant causal similarity. Hence, where it applies, the probability we assign to the disjunction of hypotheses that posit relevant causal dissimilarity must be correspondingly low. For $H(CH_2)_4COONa$ to be a kind of PD, we must be confident that all $H(CH_2)_4COONa$ is causally similar in regard of factors relevant to color of burning, and hence, the disjunction of hypotheses that posit causal dependence of color of burning on any factor more specific than *being $H(CH_2)_4COONa$* must be low.

What about priors for hypotheses that posit relevant causal similarity, “being $H(CH_2)_4COONa$ or something invariantly associated with that causes yellow burning”, “being $H(CH_2)_4COONa$ or something invariantly associated with that causes green burning” and so forth? While (RCS) in tandem with observations of yellow burning $H(CH_2)_4COONa$ allows us to well confirm “being $H(CH_2)_4COONa$ or something invariantly associated with that causes yellow burning”, it doesn’t directly constrain its prior. We might be tempted to apply a principle of indifference here; if we’ve no reason to favor any color of burning over another, perhaps we should evenly divide our confidence in relevant causal similarity between these hypotheses. If we give 0.9 to the probability that all $H(CH_2)_4COONa$ is relevantly causally similar and distinguish 9 shades, then we assign 0.1 to each. Michael Huemer (2009) has proposed a principle of indifference that invokes explanatory considerations, and Ted Poston (2014) also endorses this approach.

Whether or not we think such principles rational, they can’t cover these kinds of cases. An agent will commonly have some background evidence that favors $H(CH_2)_4COONa$ burning some color over others e.g., an agent who has never observed $H(CH_2)_4COONa$ burn but has observed some other kind of sodium salt burn yellow and no non-yellow burning sodium salts, has reason to favor yellow burning even if only to a marginal degree. Indeed, there may be many weak, analogical pressures on such priors, and as witnessed by Darwin’s and Huygens’s use of analogy, rational scientists are responsive to such pressures.

It’s hard to believe these pressures support some rationality principle that dictates precise values for these priors. However, we can remain agnostic about that. As we shall now see, existing features of our account dispose of the problem.

We'll first specify the likelihoods. Let e_i specify that a $H(CH_2)_4COONa$ sample burns the i^{th} color, and h_j be "being $H(CH_2)_4COONa$ or something invariantly associated with that causes burning with the j^{th} color". Hence, for any P(.), $P(e_i|h_j) = 1$ for i = j, and 0 for $i \neq j$. Suppose our first sample burns yellow, say, and let's designate yellow as the third color. At that point, we immediately decisively disconfirm all the uniform color hypotheses, bar h_3, since for all $i \neq 3$, $P(e_3|h_i) = 0$. Thus, *given that observation of yellow burning doesn't modify our confidence in relevant causal similarity*, we should assign the entire 0.9 to h_3, i.e., our confidence in the one surviving hypothesis that posits relevant causal similarity will equal our (prior) confidence in relevant causal similarity, $P(h_3|e_3) = 0.9$.

What priors yield this outcome? Given the specified likelihoods, preserving confidence in relevant causal similarity demands that the proportion of the probability we assign to relevant causal similarity that initially gets assigned to h_i must equal our expectation that the next sample burns the i^{th} color. So, if P(all $H(CH_2)_4COONa$ is relevantly causally similar) = 0.9, and the probability of the next sample burning yellow is, say, $P(e_3) = 1/9$, $P(h_3) = 1/9 \cdot 0.9 = 0.1$. The general requirement is this:

> Matching Constraint: The proportion of the probability assigned to relevant causal similarity (i.e. the disjunction of hypotheses that posit invariant color of burning for all sodium fluoride) that is assigned to $P(h_i)$ should equal $P(e_i)$ i.e., $P(e_i) = P(h_i)$ / P(NaF relevantly causally similar).

Given Matching, Bayes's theorem indeed dictates that $P(h_i|e_i) = 1 / [\ P(h_i) / P(\text{NaF relevantly causally similar})]\ .\ P(h_i) = P(\text{NaF relevantly causally similar})$.

So, we don't need a rule that assigns priors to the h_i hypotheses. Matching, plus the specified likelihoods, dictates their posterior probabilities. Moreover, assigning precise values to $P(h_i)$ and $P(e_i)$ is of merely academic interest. Provided we observe Matching, the value of the prior, $P(h_i)$, is immediately washed out by our first observation.

As stated earlier, learning data such as e_i shouldn't undermine our confidence in relevant causal similarity. Given reasonable background commitments, this is just inductive rationality. We can now give a more precise characterization of those commitments.

A rational scientist will not ignore any specific factor they suspect might be causally relevant to color of burning: if they suspected the relevance of some specific factor whose relevance would conflict with relevant causal similarity, that's something they would have to check out and effectively eliminate. Hence, our scientist will not merely have justified high confidence in relevant causal similarity: they will have done as much as they reasonably can to elevate that confidence by *eliminating any such specific factors from contention*.[10] Let's call this the Elimination Requirement (ER). For an agent who satisfies (ER), to the extent that their confidence in relevant causal similarity deviates from 1, it must be entirely, or almost entirely, due to a general humility in the face of possible *unknown* factors that might be relevant. Factors they can't even specify, because they're unknown, will generally provide no reason to favor a given outcome on this particular trial as opposed to any other i.e., $P(e_i|\text{there are additional}$

[10] Incidentally, this need not involve antecedently observing that cases of that kind burn yellow. For instance, we might take it that chemical kinds specify the kinds of PDs on general grounds, or if not, that "drilling down" to the level of different crystal structures of chemical kinds specifies the kinds of PDs, again on general grounds.

factors incompatible with relevant causal similarity) = $P(e_i)$ for each i. Hence, observing green or yellow, say, on the first trial will not confirm the presence of such unknown factors and diminish our confidence in relevant causal similarity. So, an agent that's rational in the specified sense will preserve that confidence when they first conditionalize on such data. [11]

Indeed, the RCSM explains why (ER) is a rationality requirement. There'd be little point in attempting to use this methodology to well confirm generalizations that presuppose relevant causal similarity if our confidence in relevant causal similarity would be severely shaken by merely conditionalizing on the salient data. Eliminating specific factors that we suspect might undermine relevant causal similarity is part of the due diligence needed to properly motivate and accomplish the task.

So, rational causal inference ultimately pulls the strings here. The only constraint we must observe for the h_i hypotheses is the Matching constraint, which is dictated by our observance of (ER). (ER) also imposes a similar matching constraint on hypotheses that posit additional dependencies that violate relevant causal similarity.[12] Beyond that, (ER) demands that explicitly formulated hypotheses from the latter group get an exceedingly low probability. However, there's no pressure to make those probabilities uniform as with a principle of indifference.

---

[11] By contrast, consider an agent with the specified likelihoods and confidence 0.9 in relevant causal similarity, for whom observing a sample burning yellow significantly confirms some hypothesis incompatible with relevant causal similarity. Suppose $P(h_3) = 0.1$ as before, but we countenance a sole non-invariant hypothesis, g, that also predicts yellow on the first observation. Since it's our sole non-invariant hypothesis, $P(g) = 1 – 0.9 = 0.1$, and hence, $P(e_3) = P(e_3 | h_3) P(h_3) + P(e_3|g)P(g) = 0.1 + 0.1 = 0.2$, whereas the probability assigned to e_3 by the Matching constraint is $0.1/0.9 = 1/9$. Consequently, by Bayes's theorem, $P(h_3|e_3) = 0.1/0.2 = 0.5$. Our new probability for h_3, which is also our new confidence in relevant causal similarity, is much less than it would be with Matching, because we've shifted confidence onto the non-invariant hypothesis, g, for which $P(g|e_i) = 0.1/0.2 = 0.5$

[12] Let D_3 be the disjunction of non-invariant hypotheses that agree on yellow being the outcome of the specified observation. So, $P(e_i|h_3) = P(e_i| D_3) = 1$ if i = 3 and 0 otherwise. By Bayes's theorem, an agent will have $P(h_3|e_3)$ = P(NaF relevantly causally similar) and $P(D_3|e_3)$ = P(NaF not relevantly causally similar) if and only if $P(e_3) = P(h_3)$ / P(NaF relevantly causally similar) = $P(D_3)$ / P(NaF not relevantly causally similar).

The resultant confirmational phenomenology is a reasonable fit with scientific practice**.** We don't typically find scientists haggling over priors for causal generalizations or the associated evidence statements, and given Matching we can see why. Further, in science we commonly require very modest numbers of observations to strongly confirm causal generalizations regarding, say, the conductivity of a kind of metal or the boiling point of some chemical kind at standard pressure. Given confidence in the reliability of our apparatus, once can be enough. Similarly, in the RCSM, given high confidence in relevant causal similarity, the first observation is quite decisive.[13]

We also have a clearer picture of the value of justified high confidence in relevant causal similarity. In tandem with (ER), it's an engine for rapid, rational confirmation of causal generalizations, a highly valuable scientific prize.

4.2 Formalizing Analogy, Consilience, and Simplicity

We've covered the justification of (1) in our Millian arguments. We'll now address the inference to the ultimate conclusion, (9), illustrating it for the second argument i.e., (9) is "*being of form $H(CH_2)_nCOONa$ or something invariantly associated with that* causes a yellow flame".

There is a kind of PD for each value of n, individuated by the belief that the molecular structure specified by that value provides a reason to suspect that cases of that kind might be disconfirmers. However, once we accept the argument, we reject those beliefs and indeed

[13] Further observations of yellow burning $H(CH_2)_4COONa$ samples may marginally boost our already high confidence in relevant causal similarity, and hence, in h_3. That high confidence allows that there might be some unknown, deviant species of $H(CH_2)_4COONa$ out there, and our probability that there is can be reduced by observations confirming that *prima facie* unremarkable samples are indeed unremarkable—repetitions are not pointless, but you only get the Nobel prize if something unexpected happens. Consequently, Matching will typically hold only to a high degree of approximation.

should assign those propositions low probabilities. Hence, our new probability distribution cannot be obtained by conditionalization.[14]

In particular, the inductively rational inference from (1) to (2) justifies rejecting the beliefs that individuate those kinds of PDs. Further, as per premise (3), the features that delineated our initial set of kinds of PDs exhausted the factors that might be causally relevant dissimilarities between salts of form $H(CH_2)_nCOONa$. Hence, given that we hold (3) highly probable—something our regulative constraint, (ER), directs us towards by demanding we eliminate every suspected factor from contention—our new distribution must assign a high probability to there being no causally relevant dissimilarities between sodium salts of form $H(CH_2)_nCOONa$, i.e., it should assign high probability to (4), "all such sodium salts are relevantly causally similar".

Let's call this new distribution $P^*(\cdot)$. Since it assigns a high probability to there being no causally relevant dissimilarities between sodium salts of that kind, and, as per (5), we hold that being a salt of form $H(CH_2)_nCOONa$ or something invariantly associated with that is a relevant causal similarity given that condition (4) is met, it satisfies (RCS) for the causal generalization C_i "being of form $H(CH_2)_nCOONa$ or something invariantly associated with that causes yellow burning". So, $P^*(C_i | e_1 \& e_2 \& ... \& e_n)$ must have a high value, for even modest values of n, where the e_j statements are now simply of the form "$S_C a$ & Ba" where S_C = *being a sodium salt of form $H(CH_2)_nCOONa$*. Since our numerous prior observations of a variety of chemical kinds of such sodium salts provided a large number of such data reports, P* should assign each of those probability 1. So, there's a more than modest conjunction of such statements for which $P^*(e_1$ &

[14] If p(A) = 1, then p(A|B) = 1 for arbitrary B, provided that P(B) > 0.

e_2 &...& e_n) = 1. Hence, in accord with (RCS), $P^*(C_i)$ must be high / well confirmed i.e., we can reasonably infer (8) and (9).

In effect, we've consolidated our initial set of kinds of PDs into one kind of case, *sodium salts of form* $H(CH_2)_nCOONa$, for which we're highly confident that all the subsumed cases are relevantly causally similar. It isn't a kind of PD, however, since we don't believe that something about that kind of case provides a reason to suspect disconfirmation. Indeed, we're highly confident nothing about those kinds of cases could cause disconfirmation.

Given the above formalization of analogy, no additions are required to formalize consilience and simplicity. A consilient hypothesis may be well confirmed by such arguments by analogy—i.e., by adopting a new distribution as specified—whereas non-consilient hypotheses can't; and a simple theory may be consilient where a non-simple one won't.

Not using conditionalization to model these powerful analogical inferences might prompt suspicion of some kind of sleight of hand. However, we made this choice because it seems to track our epistemic evolution: we initially believe it possible that certain kinds of cases could cause disconfirmation and subsequently don't, and conditionalization can't model that transition. The relationship between beliefs and degrees of belief is a fascinating area of contemporary research and we're not hostile to the possibility of alternative formal models. However, *prima facie*, this approach is a good fit.

We didn't specify a rule that uniquely determines the new distribution. There may not be one. What we will defend is that the agent's rational evolution is constrained as above. Introducing a new distribution that demotes some background beliefs could expose us to a diachronic Dutch book, but what of it? Once we acknowledge the need to sometimes discard beliefs, we must reject conditionalization as a general policy for updating. Further, as Douven (2013, 430) and others have observed, there's no inconsistency in holding conflicting views on the fairness of bets at different times. Such diachronic conflicts result from Bayesian conditionalization itself. So, even if Dutch Book arguments specify synchronic rationality constraints, there's no pressure to generalize to the diachronic case.

5. Placing the Account: Loveliness, the Voltaire Objection, and the Bad Lot

In section 6, we'll generalize our account. However, that won't affect its relation to Bayesianism, which we'll discuss now. It's manifestly not a heuristic view. Our objective Bayesian rationality constraints are explicitly motivated by causal / explanatory considerations. It also, therefore, differs from Henderson's (2014, 2017) emergent compatibilism on which (2017, 248) "a Bayesian model of the preference for better explanations can be provided, making only assumptions which are … independent of any explanatory concerns."

Jonathan Weisberg (2009, 137-141) has suggested three ways objective Bayesians could incorporate explanatory considerations: piecemeal; assigning probabilities to worlds based on how neatly their matters of particular fact can be systematized, along the lines of Lewis's Best System Analysis of laws, say; and using explanatorily driven principles of indifference. Causal inference underwrites a unified approach to multiple explanatory virtues. So, our view isn't piecemeal. Nor does it appeal to systematisability or laws. Indeed, as Ward (forthcoming) argues, we can plausibly accommodate law confirmation as a species of causal confirmation.

Finally, as seen, it doesn't appeal to principles of indifference. So, we have a distinctive objective Bayesian model of IBE.

How does it treat explanatory loveliness? As Lipton observes, if we're not to trivialize IBE, the best explanation cannot merely be the likeliest; it must be the loveliest, the one (2004, 59) "which would, if correct, be the most explanatory or provide the most understanding". However, favoring loveliness immediately suggests the Voltaire objection (2004, 143): "we are to infer that … the loveliest explanation of our evidence is, therefore, the explanation that is likeliest to be true. But why…believe that we inhabit the loveliest of all possible worlds?"

Clearly, we don't trivialize IBE. Consilience and simplicity contribute to loveliness, and the data can justify an argument by analogy that well confirms a consilient hypothesis but not its non-consilient competitor, and lack of simplicity impedes confirmation because it reduces consilience. However, it's no part of our confirmational story that lovelier hypotheses gets higher priors. Lovelier theories get better confirmed by suitable data, not because of a prior prejudice in favor of lovelier worlds, but because they're supported by *inductively rational* arguments by analogy. Hence, as applied to our account, the Voltaire objection reduces to a form of inductive skepticism.

As a bonus, we also have a clear response to van Fraassen's (1989, 143) Bad Lot objection, that IBE can't be a good rule for the formation of expectations:

> "…for it is a rule that selects the best among the historically given hypotheses.
>
> We can watch no contest of the theories we have so painfully struggled to

> formulate, with those no one has proposed. So our selection may well be the best of a bad lot"

On our account, IBE doesn't automatically well confirm / select the best available hypothesis. A sufficiently unlovely theory will never be selected, whether or not it's the best available: we can't even gather the data required to well confirm the hypothesis that assigns a distinct explanation to the yellow burning of each of an unlimited set of kinds of sodium salts, since its lack of consilience precludes confirmation by analogy. Moreover, well confirming any highly general theory's application across the variety of subsumed cases requires data that support the requisite arguments from analogy—it must be both lovely and strongly supported by the data. Being the best of a bad lot won't suffice.

6. Generalizing the Account: Explanatory Virtues and Competing Paradigms

Our arguments from analogy justify strongly confirming a consilient generalization only if we have sufficient confidence in its premises, notably those that presuppose the observed sodium salts realize the specified molecular structures. Similarly, for (RCS) to license well confirming a causal generalization regarding the behavior caused by having a particular molecular structure, conditionalization on the data must yield confidence 1 that the relevant kind of molecule is realized in the observed cases. These conditions are satisfied, at least to a reasonable approximation, in many contexts. Chemists commonly *believe* that matter is composed of molecules.[15]

However, when competing theories posit radically different mechanisms, we commonly won't have such beliefs. An early nineteenth century physicist might reasonably neither believe that light is a wave nor that it's a particle. This doesn't render the apparatus of the RCSM irrelevant. In such contexts, it characterizes rational inference *conditional on the assumptions of the relevant paradigm*. Such assumptions will specify in broad terms, the relevant mechanism, its behavior, and the connection between it and the phenomena to be explained. Thus, the assumptions of that wave paradigm would include the posit that light consists of waves, a rule for the propagation of waves along the lines of Huygens' principle, a general outline of the rules for constructive and destructive interference, and propositions specifying how light is realized by waves. These assumptions provide a deductive connection from the data to particular instances of causal generalizations regarding waves. Conditional on them, measurements of angles of incidence and refraction can, for instance, well confirm the wave mechanical generalization "passing from air to quartz causes the incident waves to propagate so that the angle of the resultant beam is less than its angle of incidence". Similarly, observations that cover a broad variety of media can justify an argument from analogy that well confirms "Light waves passing from a medium to a denser medium cause the angle of the refracted beam to be less than that of the incident beam", again, conditional on the assumptions of the wave paradigm. This use of the paradigm's assumptions strongly resembles the theory-dependent aspect of

[15] For some of the relevant evidence, see Salmon's (1984, 213-227) account of the variety of experimental determinations of Avogadro's number.

Glymour's (1980) instantiationism, although the role we give to causal considerations distinguishes it from that syntactically driven theory.

We can formally characterize *confirmation relative to a paradigm* by (a) introducing the notion of a Paradigm Dependent kind of PD (PDPD), where high confidence in relevant causal similarity and belief that cases of that kind provide a reason to suspect disconfirmation are now conditional on the assumptions of the paradigm, (b) converting (RCS) into (RCS-Paradigm) by replacing background beliefs, K, with their conjunction with the paradigm's assumptions throughout, and (c) recognizing arguments from analogy as justified relative to a paradigm if their premises are justified granted those assumptions. We'll not detail these modifications here. Their net effect is straightforward: confirmation relative to a paradigm works as confirmation *simpliciter* did when one mechanism was believed. We'll now briefly explore some broad features of this more general account.

Let r_W be some hypothesis formulated within the W-paradigm, and W be the assumptions of that paradigm. We can write $P(r_W) = P(r_W | W).P(W) + P(r_W | -W).P(-W)$. A hypothesis made from within a particular paradigm will typically be held incompatible with denial of that paradigm. So, we'll commonly have $P(r_W | -W) = 0$, and hence, $P(r_W) = P(r_W | W).P(W)$. Thus, the evaluation of r_W resolves into (a) an internal component, $P(r_W | W)$, its degree of confirmation conditional on the assumptions of the paradigm, and (b) an external component, P(W), our confidence that the posited paradigm is correct. Since the explanatory virtues assert themselves via the availability or unavailability of paradigm dependent arguments by analogy, they're primarily relevant to the internal component. Other things being equal, lovely theories can be well confirmed relative to their own paradigm whereas unlovely ones can't even achieve that.

The paradigm dependent character of the internal component suggests parallels with Kuhn's (1970, chapters II-III) account of normal science. However, the external component allows us to say quite un-Kuhnian things about confirmation *simpliciter*. Granted the consilience of the Copernican explanation of the retrogression of the superior planets, r_W, and its predictive success conditional on the assumptions of that paradigm, W, we can well confirm $P(r_W \mid W)$.[16] However, we might have good reason to give P(W) a low value: a rational preference for Aristotelian physics over whatever unknown dynamics would be required to underpin the lunacy of an Earth in motion. More generally, our evaluation of competing paradigms and of whether the paradigms under consideration are likely to include the true one may rationally inform the external component. There's no pressure to embrace incommensurability.

We can't further develop this picture here. However, as it stands, it provides a response to John Norton's (2021, chapter 9) recent objection to IBE. Norton argues that explanatory considerations are never decisive in theory choice, because putative cases of IBE are generally cases where one theory proves adequate to the evidence and the competitors simply fail—they're refuted or incur evidential debts they can't discharge. He concludes that IBE makes no substantive contribution to our understanding of confirmation.

Norton's premise is *prima facie* plausible. Observations often decisively put paid to one paradigm and at least temporarily vindicate another. However, on our model, there's good reason to reject his conclusion: explanatory virtues are required for a theory to even be a contender. Unlovely theories, even if empirically adequate to the data, will have a negligible

[16] We're assuming r_W's consilience for present purposes. We can't provide a detailed treatment here.

internal component of confirmation. So, even if the last move in the game is a decisive refutation of one competitor, it only got to be a contender because it's sufficiently lovely. Indeed, the unrefuted theory wins—as opposed to being the best of a bad lot—only if it provides a sufficiently lovely explanation of the data. So, IBE has an important role.

We might also object to Norton's premise. Taking a leaf from Kuhn's book, what looks like refutation may be terminal explanatory unloveliness. An accumulation of auxiliary hypotheses needed to salvage a declining research program's consistency with the data compromises its simplicity, precluding its current version from being internally strongly confirmed, and giving inductive grounds for pessimism regarding the loveliness of any future versions.

7. Conclusion

We have a strong case that causal inference is, at least in part, IBE. Focusing on our development of Mill's method of agreement, we obtained an account of the confirmation of causal generalizations on which mechanism, analogy, consilience, and simplicity all have significant confirmational roles. The resultant confirmation theory provides a novel approach to reconciling Bayesianism and IBE.

There are considerable opportunities for further research here: in philosophy of science, exploring the role of analogy and the associated explanatory virtues in the history and practice of science; in causal inference, extending this approach to cover Mill's methods of difference and concomitant variation. Moreover, causal inference is a vibrant area of contemporary research. So, these may be first steps in the development of a much richer IBE research program.

References

Bird, A. (2017). "Inference to the Best Explanation, Bayesianism, and Knowledge" in K McCain and T. Poston (eds.) *New Essays on Inference to the Best Explanation*, Oxford: Oxford University Press., 97-120.

Darwin, C. (1962). *The Origin of Species*. New York: Collier.

Dellsén, F., (2018). "The Heuristic Conception of Inference to the Best Explanation," *Philosophical Studies*, 175, 1745–1766.

Douven I. (2013). "Inference to the Best Explanation, Dutch Books, and Inaccuracy Minimisation". *Philosophical Quarterly* 63, 428-444.

Glymour, C.N., (1980). *Theory and Evidence*. Princeton: Princeton Univeristy Press.

Henderson, L. (2014). "Bayesianism and Inference to the Best Explanation", *British Journal for the Philosophy of Science*, 65:687–715.

Henderson, L. (2017), "Bayesianism and IBE: The Case of Individual vs. Group Selection" in K McCain and T. Poston (eds.) *New Essays on Inference to the Best Explanation*, Oxford: Oxford University Press., 248-261.

Huemer, M. (2009). "Explanationist Aid for the Theory of Inductive Logic", *British Journal for the Philosophy of Science*, 60, 345-375.

Huygens, C. (1950). *Treatise on Light* (trans. Silvanus P. Thompson). Chicago: University of Chicago Press.

Lipton, P. (1991) *Inference to the Best Explanation*. New York: Routledge.

Lipton, P., (2004). *Inference to the Best Explanation* (2^{nd} Edition). New York: Routledge.

Kuhn, T.S. (1970). *The Structure of Scientific Revolutions* (second edition, enlarged). Chicago: The University of Chicago Press.

McGrew, T. (2003). "Confirmation, Heuristics, and Explanatory Reasoning" *The British Journal for the Philosophy of Science*, 54, 553-567.

Mill, J.S. (1882). *A System of Logic, Ratiocincative and Inductive (8^{th} edition)*. New York: Harper and Brothers.

Norton, J.D. (2021). *The Material Theory of Induction*. Calgary: Univeristy of Calgary Press.

Okasha, S., (2000). "van Fraassen's Critique of Inference to the Best Explanation", *Studies in the History and Philosophy of Science,* 31, 691- 710.

Poston, T. (2014). *Reason and Explanation*, Basingstoke: Palgrave Macmillan.

Rappaport, S. (1996). Inference to the Best Explanation: Is It Really Different from Mill's Methods? *Philosophy of Science*, 63(1):65–80.

Salmon, W. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

Thagard, P. (1978). "The Best Explanation: Criteria for Theory Choice". *The Journal of Philosophy* 75, 76-92.

Thagard, P. (2021). "Naturalizing Logic: How Knowledge of Mechanisms Enhances Inductive Inference". *Philosophies* 6: 1-13.

van Fraassen, B.C. (1989). *Laws and Symmetry*. (New York: Oxford University Press).

Ward, B. (2022). “Resolving the Raven Paradox: Simple Random Sampling, Stratified Random Sampling, and inference to the Best Explanation”. *Philosophy of Science* 89, 360-77.

Ward, B. (forthcoming). “Why Confirm Laws?”. *British Journal for the Philosophy of Science*.

Weisberg, J. (2009). “Locating IBE in the Bayesian Framework”. *Synthese* 167, 125-143.

Whewell, W. (1847) *The Philosophy of the Inductive Sciences, Founded Upon Their History*, Vol. 2. London: John W. Parker.