Discovery without a 'logic' would be a miracle

Benjamin C. Jantzen

Abstract Scientists routinely solve the problem of supplementing one's store of variables with new theoretical posits that can explain the previously inexplicable. The banality of success at this task obscures a remarkable fact. Generating hypotheses that contain novel variables and accurately project over a limited amount of additional data is so difficult—the space of possibilities so vast—that succeeding through guesswork is overwhelmingly unlikely despite a very large number of attempts. And yet scientists do generate hypotheses of this sort in very few tries. I argue that this poses a dilemma: either the long history of scientific success is a miracle, or there exists at least one method or algorithm for generating novel hypotheses with at least limited projectibility on the basis of what's available to the scientist at a time, namely a set of observations, the history of past conjectures, and some prior theoretical commitments. In other words, either ordinary scientific success is miraculous or there exists a logic of discovery at the heart of actual scientific method.

Keywords logic of discovery, scientific discovery, confirmation, induction, machine learning

1 Introduction

Deductive logic has two aspects. On one face is the relation of logical consequence, the guarantee of truth provided for one assertion by the truth of one or more others. On the other face are methods of production. Valid inference schemes, such as modus ponens or proof by cases, may be used to generate truths from truths. These inference schemes can be formalized in systems of 'natural deduction' such that true statements can be mechanically produced from true statements by chaining the rules of the system.

Inductive logic, Humean worries notwithstanding, is widely supposed to involve a relation of support analogous to that of logical consequence. Some statements support or increase the probability of the truth of others, though they do so imperfectly. But what of production? Is there an inductive analog to the mechanical process of deductive derivation? This question takes on special urgency in the context of the natural sciences. There, a mechanical process of inductive derivation would have to involve something like rules of conjecture or hypothesis generation. Such rules would constitute a 'logic of discovery'.

The notion that there exists a logic of discovery worth the attention of those concerned with the normative aspects of scientific methodology has fared poorly over the last century. This is due largely to the inverse success of the distinction between the contexts of discovery and justification, a legacy of logical empiricism

Benjamin C. Jantzen
Philosophy Department
229 Major Williams Hall (0126)
220 Stanger Street
Virginia Tech
Blacksburg, VA 24061 U.S.A.
Tel.: +1-540-231-1766
Fax: +1-540-231-6367
E-mail: bjantzen@vt.edu

that is traditionally traced back to Hans Reichenbach (1938). The distinction separates the production of scientific knowledge into two logical[1] phases. On the one hand, there is the process by which hypotheses are brought to stand in the proper relation with empirical facts and background knowledge so as to be justified. This relation of justification carries all of the epistemic weight for securing scientific knowledge, and is amenable to normative evaluation. On the other hand, there is the process of 'discovery' by which new hypotheses enter the stage to be vetted. This process is presumed to involve unanalyzable or irrational elements.

Explicit appeal to the discovery/justification distinction has fallen out of vogue, and the handful of arguments purporting to establish the irrationality and philosophical irrelevance of hypothesis generation have withered under scrutiny (see Section 2). Nonetheless, discussions of normative methodology continue to be dominated by a concern for post-hoc justification. While a more nuanced view of the relation between experiment and hypothesis has developed, there remains little interest in identifying and evaluating the methods by which hypotheses are produced. Why is this the case? While many concede the bare possibility of a logic of discovery, there seems to be a shared sense that it couldn't be relevant to the way in which scientists actually do their work, or perhaps more generally, for the way agents like us are able to reason inductively. Whether or not such an assumption explains the dearth of interest in the logic of discovery, the importance of that pursuit clearly turns on the truth of this claim. In what follows, I argue that general features of the history and practice of science imply that not only is a logic of discovery possible, but the actual practice of science must be driven in large measure by just such a method. The traditional view takes for granted that sufficient hypotheses are generated for science to follow a two-step dance of generate (in the context of discovery) and test (in the context of justification). To the contrary, the empirical facts of scientific success imply that generate and test cannot be good enough unless there is a non-trivial method to the generation. In fact, the method of hypothesis generation – the implicit logic of discovery – is doing so much of the epistemic heavy-lifting that post-hoc testing amounts to tidying up a fait accompli.

1.1 What is a logic of discovery?

An immediate clarification is in order. 'Logic of discovery' is one of those mercurial terms in philosophy that changes its complexion for every author that wears it. What I have in mind has been expressed best by those who most strenuously reject its existence. Hempel, for instance, considered and rejected the possibility of "...a mechanical procedure analogous to the familiar routine for the multiplication of integers, which leads, in a finite number of predetermined and mechanically performable steps, to the corresponding product"(1966, p14). The product in question is a significant scientific hypothesis. Curd (1980, p205) says that a logic of discovery is "The specification of a procedure, possibly algorithmic, that will generate non-trivial hypotheses."[2] And Laudan views "...the logic of discovery as a set of rules or principles according to which new discoveries can be *generated*"(1980, p174)(emphasis in original). As I will use the term, a *logic of discovery* is a computable method or procedure for generating one or more hypotheses from a set of empirical facts, preexisting theories, and theoretical constraints, where each hypothesis generated is significant in that it is both consistent with existing data and likely to be projectible over a substantial range of data not previously in evidence. In other words, a hypothesis is significant just if it is likely to make true predictions about previously unknown cases in its intended domain. By "likely," I just mean that the production of hypotheses by the logic exhibits a high relative frequency of successful hypotheses, and by "substantial range of data" I mean such as would be acquired in the usual practice of science.

The definition of logic of discovery I've adopted is focused squarely on the *generation* of hypotheses. For this reason, much of the revived philosophical interest in 'discovery' that concerns itself with preliminary appraisal of hypotheses, or with a "context of pursuit" between discovery and justification (Laudan, 1980) is orthogonal to the discussion here. While rational reconstructions of the choices scientists make about which hypotheses to pursue are important, such reconstructions presume that we already have hypotheses to consider. I am strictly concerned with whether or not methods for generating hypotheses exist in the

---

[1] It is often ambiguous whether the stages are supposed to be temporal or logical. See, e.g., (Hoyningen-Huene, 2006).

[2] This is actually one of three characterizations Curd considers. He deems this the least plausible of the three.

first place.[3] Similarly, the large and growing literature on abductive reasoning is not relevant. The general schema of abduction notes that hypothesis H would explain or make highly probable surprising fact E, and then infers H from E. Whatever the merits of abduction and regardless of how central a role it plays in intelligent reasoning in humans or machines, it too presupposes an explicit collection of hypotheses.[4] What is of interest to the logic of discovery is where that collection of hypotheses comes from, and how it is characterized. Backward-looking accounts of 'discoverability' that consider whether one could have rationally inferred the correct theory "had we known then what we know now" (Nickles, 1984, p20) are similarly beside the point. My focus here is on whether there are methods for generating good hypotheses looking forward from what we know now.

As my definition suggests, I understand limited projectibility to be a necessary condition for a hypothesis to be significant or to constitute a discovery. Note that I am *not* insisting that an assessment or test of limited projectibility be part of a logic of discovery. To be the sort of logic of discovery I have in mind, a procedure must be likely to produce hypotheses that have the property of limited projecibility. Such a procedure does not test or otherwise assess whether a hypothesis is in fact projectible – to do so would require examining data not used in the production of the hypothesis, and thus could not be part of a procedure of hypothesis generation! What I have in mind is closely analogous to conditions placed on confidence interval estimators in statistics or algorithms in Probably Approximately Correct (PAC) learning. In the former case, a confidence interval estimator – a method for generating hypotheses about the value of parameters in an unobserved distribution based on samples drawn randomly from that distribution – is good just if the procedure yields an interval containing the true value with high probability. In the latter case, a condition on any acceptable learning algorithm for a given problem – such as finding a function that correctly classifies objects based on their observed properties – is that it produce with high probability a function that is at least approximately correct for the class of objects in question. In neither case do the procedures for generating answers appeal to tests of the correctness or projectibility of candidate hypotheses. Similarly, I want to insist that a necessary condition for any algorithm to constitute a logic of discovery is that, with high probability, it produce hypotheses with at least limited projectibility.

Of course, you might think that limited projectibility is not a sufficient condition; we often demand that a hypothesis be explanatory or possess some other theoretical virtues before we dub it a significant discovery. That may be the case, but for my purposes, it's safe to set aside the problem of specifying a full set of necessary and sufficient conditions. That's because additional restrictions on the notion of a discovery only help my case in the sense that positing additional restrictions just makes it harder to succeed at discovery.

The given definition also leaves open the form and content of a logic of discovery. Such a logic need not resemble simple, 'enumerative induction', e.g., every raven we've seen has been black, therefore hypothesize that all ravens are black. In fact, enumerative induction cannot be the whole story. Nor need the method be elegantly axiomatizable in some first-order language. The algorithm may involve complex mathematical constructs or manipulations that although they may in principle be grounded in the axioms of set theory, would appear ghastly if cast in such a form. Of course, this is no different than most significant algorithms such as the Cooley-Tukey Fast Fourier Transform. All that is required of a logic of discovery is that the procedure be computable. That is, given finite data and finite time, the procedure will present a hypothesis likely to project successfully over further data (if any exist), and it will do so in a manner that can be captured by some step-by-step procedure or algorithm.

This is a stronger notion of 'logic of discovery' than even the 'friends of discovery'[5] typically endorse. For instance, Kelly (1987, p435) describes a logic of discovery as a procedure "guaranteed (or at least likely) to produce hypotheses that are reasonable with respect to given evidence."[6] The sort of procedure I mean to

---

[3] An interesting exception is work by philosophers on procedures for generating mechanistic hypotheses in the life sciences (Darden, 2002; Craver and Darden, 2013).

[4] That isn't quite fair if one understands abduction in Peirce's sense. Peirce took 'guessing', the inception of a new idea, to be a stage of abduction (Tschaepe, 2014). However, he seems to have thought that guessing is not rationally analyzable. Those who have touted abduction as the logic of 'discovery' (e.g., (Hanson, 1961)) often ignore the actual stage of guessing, or assume a trivial mechanism of generation, as in (King et al, 2004).

[5] According to (Savary, 1995), this felicitous phrase was coined by Gutting (1980).

[6] Kelly (1987) actually refers to such a method as a "generation procedure" and reserves the singular term "the logic of discovery" to refer to the study of all such procedures. My references to *a* logic of discovery are equivalent to Kelly's talk of "a generation procedure."

refer to is one which will produce hypotheses that are not only reasonable with respect to existing evidence, but which are likely to remain so with respect to some significant amount of future evidence.

Both the existence and the epistemic relevance of a logic of discovery of even the weaker variety have been repeatedly denied. To the contrary, I argue that there must be a logic of discovery in the strong sense if the history of scientific success is to be anything but a miracle. Articulating that logic would, it seems, be profoundly relevant to the epistemology of science.

To make my case, I will first clear a little ground by defusing a litany of common objections to the possibility and relevance of a logic of discovery. In Section 3, I present modern machine learning algorithms as a prima facie proof by example of the existence of a logic of discovery, and consider a reply from foes of discovery asserting that such algorithms are incapable of identifying and making use of previously unconceived theoretical variables. In Section 4, I clarify this problem of identifying novel variables. Finally, in Section 5 I argue that, rather than precluding a logic of discovery, the problem of novel variable identification entails a dilemma: either the history of science is one long chain of miracles, or there exists a logic of discovery that is in fact approximated by actual scientific practice. Both this claim and its supporting argument have, until now, gone unacknowledged.

2 The foes of discovery

It has been affirmed many times that there is no logic of discovery, and that even if there were, it wouldn't matter. The arguments behind these affirmations tend to be cursory (as if the idea hardly warrants attention), and as a rule they fail because they each conflate questions about logics of discovery with one or more distinct issues. My aim is to provide a positive argument for the existence of such a logic and its de facto importance to actual science. As such it is not necessary to rebut every criticism that has ever been leveled at the possibility. Nonetheless, it will be instructive to survey the arguments of the opposition if for no other reason than to forestall some objections.

One of the most tenacious tropes tendered by foes of discovery conflates normative theories of discovery with descriptive theories of the psychology of discovery, then declares the logic of discovery philosophically irrelevant because descriptive theories are the business of psychologists (Popper, 2002; Laudan, 1981). Such arguments are at best question-begging, merely defining away an otherwise substantive position. Those interested in discovery clearly do not intend to restrict their attention to *actual* human behaviors of hypothesis generation, and the study of adequate procedures of hypothesis generation is plainly normative, whether or not it is productive (Kelly, 1987). As McLaughlin McLaughlin (1982, p200) says, the normative study of logics of discovery deserves to be judged on "less whimsical grounds."

An equally superficial dismissal of the possibility of a logic of discovery presumes that any such method must not only generate significant hypotheses, but also explicitly resolve the problem of induction. Curd (1980, p207), for instance, argues for the impossibility of a logic of discovery on the grounds that theories are always underdetermined by the evidence, and thus no matter how well a generated hypothesis fits the current data, we cannot be sure that it is true. This line of argument conflates two distinct issues, and in so doing unduly burdens the friend of discovery. On the one hand, there is the justification of commitment to a hypothesis on the basis of an argument schema deemed strong. One such schema is to compute a measure of confirmation (e.g., a Bayesian posterior probability) for a given hypothesis, h, and collection of observations. If the result is above some threshold, we are said to be justified in committing to h. Alternatively, one might show that h is the output of a reliable procedure of hypothesis generation once the observations are provided as input. This, too, would serve to justify commitment to h. On the other hand, there is the problem of justifying the argument schema itself. Why is this schema strong? What reasons do we have to suppose that it reliably sanctions hypotheses that are likely to be true? A logic of discovery is a norm for hypothesis generation. It is not intended to guarantee its own reliability, any more than an explication of confirmation contains an explicit argument for the tendency of confirmation to track truth. To insist that it do so is to confuse the justification of hypotheses with the justification of a method for their generation.

A more serious critique grants that a normative theory of discovery is possible, but argues that it is irrelevant (Laudan, 1980). According to this line of thought, the only pertinent epistemic goal for the philosophical scrutiny of scientific methodology is the justification of hypotheses. Any logic of discovery would

be dominated by other methods of justification. If we use the term "generationist"[7] to denote methods of hypothesis justification that refer only to the process by which a hypothesis was produce, and "consequentialist" (Laudan, 1980; Nickles, 1985) to denote methods that evaluate an hypothesis strictly in terms of the empirical adequacy of its logical consequences, then the claim is that the warrant for accepting the truth of an hypothesis provided by consequentialist methods strictly dominates that provided by generationist considerations. [8] The argument turns on the supposition that Whewell, Herschel, and Comte were right to insist that the evidential weight of data utilized in generating a theory must be discounted relative to data which were not considered at the time of hypothesis creation. The generationist justification of a hypothesis can only appeal to its pedigree—if the hypothesis was generated in the right way, i.e., via a proper logic of discovery, then it is justified. Since only the data used to generate a new hypothesis determines its justification, data that was not considered at the time of hypothesis generation cannot play a role in its assessment. It's impossible for a generative method to make use of the good stuff for justifying hypotheses since the most compelling data can only be introduced after the hypothesis is in hand. The consequentialist ostensibly faces no such restriction. Thus, even if the generationist can identify a logic of discovery, it is at best redundant with respect to justifying hypotheses.

This argument is unconvincing because it saddles any logic of discovery with confirmationist criteria of justification. Even if we agree that evidence considered only after the formulation of a hypothesis is the most important for justification by confirmation, there's no reason to suppose this is true for justification by generationist methods. The information ignored from a confirmationist perspective—namely the way in which a hypothesis was generated—may in fact confer greater justification than any purely post hoc assessment. I'm not saying that it does, nor do I intend to endorse the claim that there is a single scale of justification along which these different methods are commensurable. My point is just that one cannot use the fact that non-consequentialist methods fail to meet confirmationist criteria as an objection to non-confirmationist methods.

Perhaps the perceived irrelevance of logics of discovery has more to do with procedures of knowledge generation rather than the relation of justification itself. Suppose that any generation procedure which can produce suitable (i.e., well-justified) hypotheses must consist of a test procedure of the consequentialist variety that is coupled to a procedure which enumerates hypotheses and applies the test. In that case, there is nothing very interesting to say about the generation procedure. All it has to do is churn out hypotheses (in any old order) and the testing procedure sorts the wheat from chaff and provides all of the justification we could need or want. As Kelly (1987) points out, however, there is a computational symmetry between test and generation procedures. Just because a method can be written as a test procedure applied to an enumeration of hypotheses doesn't mean it has to be. Test procedures can be used to build hypothesis generators, and vice versa. In many cases, it is the generation procedure that does all of the work. Kelly offers as a trivial illustration of this point the problem of finding a number y that, given any particular number x, satisfies $y = x^2$. One could imagine devising a test to be used along with an enumeration of the possible values of y. For each number spit out by the otherwise rather dumb enumerator, we would apply our sophisticated test, and stop searching whenever the test says we've confirmed a value of y. But that would be a ridiculous way to proceed in this case. It's far more efficient to simply compute the value of x^2 using the sort of algorithm we all learned in elementary school, the sort of thing you do with pencil and paper. Of course, Kelly is interested in making a non-trivial point about methods of hypothesis generation. In general, we're seeking hypotheses that satisfy one or more conditions, such as being well-confirmed or explanatory, or satisfying particular human interests, and so on. Call the relation that a good hypothesis bears to all relevant considerations a suitability relation. When we test a hypothesis, we are assessing whether or not it satisfies some given suitability relation. Kelly proves that for any suitability relation that is recursively enumerable (i.e., it is the domain of a recursive function), there exists a hypothesis generation procedure that never fails to output a suitable hypothesis if one exists, nor ever outputs an unsuitable hypothesis. In other words, if it's possible

---

[7] My use of this term is a slightly unfaithful replication of Kelly's (1987).

[8] Note that Laudan considers only the confirmationist subset of consequentialist approaches to theory justification. Falsificationist theories are thus ruled out of consideration.

to test hypotheses in some complete but arbitrary enumeration as the consequentialist envisions, then it is also possible to build a generation procedure that computes a suitable hypothesis.[9]

Of course, neither of these responses to the argument from irrelevance establishes that there is *in fact* any interesting generative procedure for science. What has been established is a bare conceptual possibility. Whatever you take a test of hypothesis suitability to accomplish, one cannot dismiss the possibility of an algorithm that directly *generates* suitable hypotheses. But one could still retreat to the position that for scientific purposes – whatever the notion of suitability is for scientific inquiry – there simply are no adequate logics of discovery, and the thus the point is purely academic. The question then is whether or not the actual aims of science admit of interesting hypothesis generation procedures.

3 Discovery algorithms are everywhere, so what's the problem?

3.1 Modern machine learning and scientific discovery

One direct way to argue that logics of discovery are not merely possible is to provide examples of non-trivial procedures for generating significant hypotheses with respect to actual scientific aims. Modern machine learning is a rich source of such examples. Freno (2009, p377) goes so far as to claim that progress in machine learning algorithms shows "...that the logic of scientific discovery is not merely a philosophical project, but it is instead a mature (and thriving) scientific discipline." Seeing why such an assessment may be warranted requires a little vocabulary. To begin with, machine learning (ML) is the study and design of computer algorithms that improve their performance at a task on the basis of their experience (Mjolsness and DeCoste, 2001). A *task* involves passing data to a *model* which then returns an output of interest. Data is described in terms of *features* – the ML term for variables. The multi-dimensional space of all possible feature value combinations is referred to as feature space, and any particular point in that space – corresponding to an actual or possible object or observation – is called an *instance*. Often, we are interested in categories to which instances belong, and these are typically referred to as *classes*. A *learning problem* consists of using training data to find a model which correctly performs a task. The principal aim of ML is the development of algorithms that solve learning problems.

To give a realistic example, a task might consist of predicting whether or not a given sequence of DNA is a gene encoding a protein. Each possible DNA sequence can be described in terms of which base pairs occupy which location, such that each location is a feature that can take on the values A, G, C, or T. The given sequence is passed as input to some model which then classifies the sequence based on its features and returns a category label, e.g., "gene" or "not gene". The associated learning problem involves using a sample of data to find a model which will correctly make this identification. Like this example, many of the tasks of concern to ML overlap with the sorts of inductive tasks relevant to empirical science. When applied to such tasks as the identification of coding sequences of DNA, ML models just are a sort of hypothesis. On the face of it, learning algorithms are logics of discovery.

In the remainder of this section, I present a number of ML models and, more to the point, the algorithms that learn them. I will not pretend to a comprehensive survey. Rather, I have chosen a few representative examples to give a sense for the variety, scope, and power of ML algorithms. To bring some order to the presentation, I group examples according to the nature of the model, largely following the three part classification scheme of (Flach, 2012). To Flach's logical, geometric, and probabilistic categories, I have added a fourth, 'analytic' grouping.

3.1.1 Logical models

Many of the models that emerged from classical artificial intelligence can be loosely described as logical or declarative because they can be translated into sets of rules that are easily understood by humans.[10] A

---

[9] The insistence on a recursively enumerable suitability relation can be dropped. It turns out that "[f]or any nonempty degree of uncomputability, there are infinitely many suitability relations...properly of that degree for which there is a strongly adequate hypothesis generation machine"(Kelly, 1987, p449).

[10] Curiously, these models seem to be the only ones to garner significant attention amongst philosophers of science, e.g., (Gillies, 1996).

logical model consists of a set of sentences in a formal language. The extension of each sentence is the subset of feature space for which the sentence is true. A logical model thus partitions the feature space into a set of cells, each of which can then be mapped to a particular sort of output depending on the class at hand. To take one simple but important example, consider the task of predictive classification. In particular, suppose we wish to determine whether or not a patient has a disease given a collection of symptoms and test results. The appropriate logical model for this task is a decision tree. At the base of the tree is feature for which there are a finite number of values. For instance, the root of the tree may contain the attribute "cough" which can be either "present" or "absent". Each possible feature value corresponds to a branch of the tree, and leads to either another node containing a feature or to a terminal 'leaf' which provides a class label, in this case either "has the disease" or "doesn't have the disease".[11] Despite their relatively ancient pedigree (a few decades is a long time in ML), decision trees remain one of the top ten models deployed in data mining (Wu et al, 2007). In scientific contexts, decision trees are used to represent significant empirical hypotheses in an increasingly wide range of fields, including bioinformatics (Darnell et al, 2007), biochemistry (Medina et al, 2013), medicine (Kenny et al, 2005), and astronomy (Way et al, 2012).

Of course, for making the point about logics of discovery, what matters is that there exist algorithms for producing the decision tree models automatically given a set of training data (Quinlan, 1992; Breiman et al, 1984). In fact, the applications cited above all make use of machine learning algorithms to generate a decision tree which represents novel hypotheses about, e.g., the subset of amino acids in a protein most prone to strong interactions with other proteins. Broadly speaking, all such algorithms proceed by selecting the feature that divides the training examples into the most homogeneous sets possible (there are a variety of scoring functions that assess the degree of homogeneity). This feature becomes the root of the tree. For each branch, the process is iterated either until some feature divides the remaining instances into sufficiently homogeneous subsets, or there are no more features to choose from.

There are other important types of logical model that represent hypotheses via formal languages and generate inferences via deduction. Rule models, for instance, are similar to tree models, though somewhat more flexible. Like trees, they represent hypotheses as sets of sentences. Unlike trees, these sentences are conditionals with an antecedent consisting of conjunctions of literals and a consequent asserting membership in one of the classes of interest. Learning algorithms for rule sets function similarly to those for decisions trees. Roughly speaking, they iteratively build up the body of a candidate rule by adding literals until all of the training examples described by that rule belong to a single class. This is repeated until all of the training examples are covered. The early algorithms for building 'expert systems', such as meta-DENDRAL (Feigenbaum and Buchanan, 1993) and INDUCE (Michalski and Chilausky, 1980) – some of the earliest in ML – are algorithms for learning rule sets.

3.1.2 Geometric models

Despite their continued popularity, tree models have retreated from the theoretical spotlight in ML. Instead, there is a flurry of development of the geometric and probabilistic models. There is great and growing diversity in these two classes such that even a concise survey must remain beyond the scope of this paper. But I can sketch a few models that are simultaneously important in ML and transparently analogous to the sorts of ordinary scientific hypotheses that are the focus of philosophical scrutiny. Let me me begin with the geometric models, so called because they involve specifying the boundaries of regions within feature space. Rather than represent hypotheses in terms of logical expressions, these models represent via shape. As with decision trees, the simplest task to which such models can be put is binary classification, and the simplest boundary is a flat surface (or a line in 2-D). A linear classifier is the specification of a plane that divides the space of possible instances in two, each corresponding to one of the two classes of interest. In effect, it is a simple theory that predicts, for any combination of observed features, whether an instance will fall in one class or the other. For example, suppose that we are interested in classifying days as rainy or dry based on other meteorological features. The obvious application here is predicting whether or not it will rain. In particular, suppose we consider just two features, namely average temperature (averaged over the day) and average humidity. These two features define a plane, with any particular day represented as a point. We could draw a line which cuts this plane in two, ideally separating all of the rainy days from the dry ones. We

---

[11] Note that any path through such a tree is equivalent to some sentence in disjunctive normal form.

could then use this model to predict, given any values of temperature and humidity, whether that day will be rainy or not. Figure 3.1.2 (a) shows a sample of actual recent weather data from my home town along with a linear classifier, a line dividing instances of rainy days (triangles) from instances of dry days (circles).
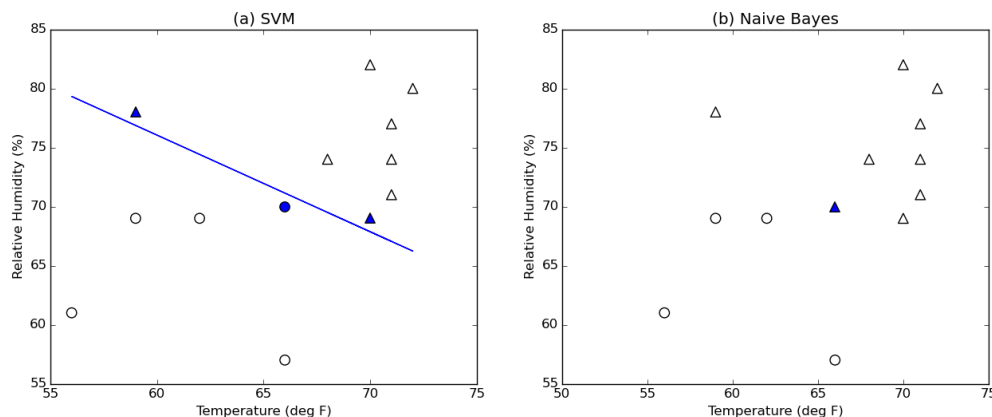

Fig. 1 (a) A decision boundary produced by a linear SVM. (b) The predictions made by a naive Bayesian classifier trained on the data shown.

But how should we select such a line? Put differently, if the line represents a hypothesis about the association between rain, temperature, and humidity, where do such hypotheses come from? One of the most powerful and important learning algorithms in ML is the Support Vector Machine (SVM). While SVMs can learn surfaces more complex than planes,[12] the simplest SVM learns a linear classifier. It does so by determining the boundary that separates positive and negative training examples while maximizing the minimum distance of the training examples to the boundary. In other words, a linear SVM divides the space of possible instances with a flat surface, so that all instances belonging to one class of interest are on one side of the boundary (if it is possible to do so), and such that the boundary is as far is it can be from known training examples in the two classes. The line drawn in Figure 3.1.2 (a) is the boundary learned by an SVM from the data shown. The filled shapes are the instances closest to the boundary. For an SVM, the learned boundary can be expressed as a linear combination of these instances. What makes SVMs so powerful is their resistance to overfitting. That is, they avoid matching the boundary too closely to the training data. Algorithms that overfit tend to err more often on new unseen data. Currently, SVMs are essential tools for hypothesis generation in bioinformatics (Yang, 2004), and are rapidly being adopted in other fields. To give one example, Han et al (2007) used an SVM to classify proteins – described in terms of continuous properties such as amino acid composition, volume, polarity, and hydrophobicity – as good or bad targets for therapeutic drug interaction.

3.1.3 Probabilistic models

Logical and geometrical models can be (and more often than not, actually are) augmented with probabilistic machinery for various tasks, such as estimating the probability that a previously unobserved instance falls within one or another class. As a rule, these models encode only the posterior probability, $P(Y|\mathbf{X})$ where Y is the target variable (e.g., the classes into which one is attempting to sort data) and $\mathbf{X}$ is a vector of features. But there is a rich set of models that encode the full joint distribution $P(Y, \mathbf{X})$. These models are 'generative' in the sense they allow one to generate new data fitting the postulated distribution. As with the other classes of model considered, these models can be applied to a wide range of tasks.

---

[12] Using 'kernel methods' that essentially apply a nonlinear transformation to the feature space to get a new space that is linearly separable.

Again, the simplest task to consider is classification, and the simplest[13] probabilistic classifier is the so-called 'naive Bayes' model. Such a model assumes that the probability of one feature taking on a given value is independent of all other features conditional on a choice of class so that $P(X_1, X_2, \ldots, X_n|Y) = P(X_1|Y)P(X_2|Y)\cdots P(X_n|Y)$. Though this assumption is seldom justified, it does not generally impair the performance of these classifiers provided the proper learning algorithm is deployed. There are a variety of such algorithms, depending on the nature of the features considered (whether they are continuous or discrete) and what assumptions one is willing to make about the distribution over classes. The algorithm for learning in the case of continuous valued features that are presumed to be normally distributed within each class is a simple procedure of statistical estimation: the sample means and variances from training instances in each class are used as estimates of the parameters in the corresponding Gaussian distributions. The result of learning a naive Bayesian model for the weather example discussed above appears in Figure 3.1.2 (b). While, unsurprisingly, it gets most of training set (the data shown) correct, note that the learned model mis-classifies the single filled shape (the filled triangle should be an open circle). Here again we have an algorithm that, irrespective of domain, produces hypotheses concerning the association of a target variable (the classification of rainy or dry in this case) and a collection of features (e.g., temperature and relative humidity). These popular classifiers have successfully generated novel and significant hypotheses in wide variety of scientific contexts. For instance, a naive Bayes classifier has been successfully applied to the problem of classifying bacteria – generally unseen and often new to science – on the basis of ribosomal RNA (Wang et al, 2007).

Another important type of model, at least from the perspective of scientific discovery, is the causal network. Causal networks represent the causal relations amongst a set of variables by way of a directed acyclic graph (DAG) with variables as nodes. In order to capture probabilistic causal relations such as that which obtains between smoking and lung cancer or high cholesterol and heart disease, the DAG is supplemented with a probability distribution. In particular, each node stores a distribution over the possible values of the corresponding variable, conditional on the values of its parents in the graph. Causal networks represent powerful predictive and explanatory hypotheses. They allow one to predict the values of a subset of variables given observations of another subset, or to foretell the consequences of intervening on the value of a variable, e.g., will the risk of heart disease go down if one intervenes to reduce blood cholesterol levels? They can also be used to infer the most likely cause of a particular outcome. And like the other models discussed above, there exist algorithms for automatically generating these causal hypotheses from data or sequences of data. Algorithmic approaches to causal discovery have a deep history going back at least as far as Francis Bacon. Modern methods (see, e.g., Pearl, 2000; Spirtes et al, 2000) rest upon a non-reductive, interventionist account of the casual relation, and use conditional probabilistic independence relations to infer causal structure. To cite just a couple of examples, algorithms in this class have been applied to the difficult problem of discerning causal relations in climate science (Ebert-Uphoff and Deng, 2012) and to the search for interactions amongst genes for the endoplasmic reticulum (an essential component of eukaryotic cells)(Battle et al, 2010).

3.1.4 Analytic models

Finally, there is a class of models that, for lack of a better term, I'll call 'analytic'. These models represent hypotheses via symbolic, analytic mathematical expressions, typically equations indicating law-like relations amongst real-valued variables. There are two principal sorts of learning algorithms used to generate these models. On the one hand, there is heuristic search (see, e.g., Simon et al, 1981; Zytkow and Simon, 1988; Langley and Zytkow, 1989). This approach treats discovery as search through a given problem space facilitated by the use of heuristic rules for choosing a direction in which to search. For example, one of the earliest learning algorithms in this class, BACON.1, searches for simple quantitative relations between two variables (see Langley et al, 1987). To do so, it attempts to find algebraic combinations of the variables that result in an invariant value. To find these combinations, it uses heuristics such as: if two variables increase together, consider their ratio. BACON.1 was able to 'rediscover' Boyle's Law and other simple empirical relations. Successors such as BACON.3 use nested search to discover relations amongst multiple dependent variables. More recently, Caschera et al (2010) have used a heuristic search involving iterated experimentation to optimize a liposomal drug delivery system. The result was a system with twice the drug packed into each fatty delivery vesicle (liposome) as the previous standard.

---

[13] The naive Bayes classifier is also one of the oldest (Maron and Kuhns, 1960).

Models like that found by the BACON lineage are equations. Another set of analytic models encode invariants of the motion as symbolic functions. An invariant of the motion is a function of the dynamical variables the value of which does not change through time. Linear momentum, angular momentum, and the Lagrangian are all invariants of the motion of systems obeying classical mechanics—they are all conserved through time. The automated discovery algorithm of Schmidt and Lipson (2009) takes time-series data and performs a search through a space of symbolic mathematical expressions, looking for non-trivial invariants of the motion. This search is made via a genetic algorithm. That is, from a seed population of candidate expressions, random variations of candidates are generated and scored with respect to how well they actually predict relations amongst first derivatives of quantities in the data. The best variants are kept, and the process is repeated until sufficiently 'fit' candidates are developed. This algorithm successfully (re)discovered the correct Lagrangian for the double-pendulum amongst other systems with complex dynamics from experimental data.

3.1.5 Summing up

I've left a lot out of my survey of ML models and associated learning algorithms. For instance, I've said nothing about the use of artificial neural networks for classification and prediction or the development of 'deep-learning' (Bengio, 2009). Nor have I considered algorithms for learning regression models. This is not because these algorithms are less important or less impressive as generators of novel empirical hypotheses, but simply because the set of examples is far too rich to compile here. And that is precisely my point. A quick glance at a contemporary textbook (e.g., Russell and Norvig, 2009; Bishop, 2007) is enough to make the case that the discovery of models or theories that successfully explain, predict, or control the world is now very much the business of machines.

3.2 The 'deep structure' objection

Given that algorithmic methods exist in such abundance, and that they have in fact produced plausible and heretofore unconceived hypotheses about empirical regularities, causal relations, or even law-like relationships between variables, where is the problem? How is this not simply a proof by example (or legion of examples) that the foes of discovery are mistaken, at least about the impossibility of a logic of discovery? Here again, Laudan and Hempel can serve as spokesmen for the opposition. According to Laudan (1981), the algorithms described above are at best suitable for establishing 'observable regularities'.[14] For such purposes, he concedes that a logic of discovery is possible, as indeed he must given the above examples. What he intends to deny is the possibility of a logic of discovery for generating hypotheses about 'deep structure'. Though no precise account of deep structure is given, Laudan says it is described by explanatory theories "...some of whose central concepts have no observational analogue" (1981, p186). Much the same assertion could be found in Hempel's work a decade earlier: "[S]cientific theories and hypotheses are usually couched in terms that do not occur at all in the description of the empirical findings on which they rest" (1966, p14). In order to produce such a theory then, a logic of discovery would have to "...provide a mechanical routine for constructing, on the basis of the given data, a hypothesis or theory stated in terms of some quite novel concepts, which are nowhere used in the description of the data themselves"(Hempel, 1966, p14). More recently, Woodward (1992, p200) has put the objection this way:

> A computer programmer armed with an arsenal of concepts and definitions can write a program that, employing random search procedures of varying sophistication, explores the various possible interrelationships of the variables contained in those concepts. "New" concepts can be generated by clustering the variables already present in different ways. But such clustering will not produce conceptual shifts like that from the concept of impetus of the middle ages to the principle of inertia— the key idea in the Scientific Revolution of the 16th and 17th centuries.

---

[14] Obviously, neither Hempel nor Laudan could have been aware of all of the methods reviewed here when they framed their objections, as many are recent inventions. But enough – such as decision trees, naive Bayes classifiers, and (for Laudan at least) the BACON lineage of heuristic search algorithms – were around by the time they penned their objections.

Without relying on an observable/unobservable distinction, one might neutrally reconstruct the challenge in the following way. Important new hypotheses are generally framed in terms of variables distinct from those with which the phenomena to be explained are described. Often, one posits the existence of a new kind of entity with new kinds of properties, and then explains the previously observed regularities by appealing to law-like relations amongst the new properties, as well as laws or bridge laws that connect them to the old observed properties. What Hempel, Laudan, Woodward, and other foes of discovery find implausible is that there could exist a method for introducing such novel variables. Put somewhat differently, if we already know the variables amongst which the laws we seek obtain, then finding those laws is a matter of curve-fitting. No one denies machines can do this. But all the *real* work of discovery is done in specifying the variables in the first place. What mattered is that Kepler chose to work in terms of distances and periods relative to the sun. The rest was curve-fitting. Once we know to look at temperature, pressure, and volume together, the rest is algebra. The hard part, then, is coming up with the quantities we should be trying to relate. Although no clear argument has been offered to this effect, the claim is that such an act of novelty cannot be captured by an algorithmic procedure.

3.3 Rebuttals from ML

The theory and practice of feature construction (variable invention) is a central concern of ML, and there is a large and growing collection of methods for learning new features. Many of these, such as Principal Component Analysis (PCA), express new features as functions of old. For this reason, they do not necessarily escape the deep structure objection (though one might reasonably argue that conceptual novelty is compatible with a functional relation between old and new variables). But there are also plenty of examples of learning algorithms which generate novel variables that are *not* merely functions or groupings of given variables. In the interest of space, I'll sketch two of these.

The first is a method for positing latent (i.e., unobserved) variables based on failures of the naive Bayes assumption (Zhang et al, 2004). As I said above, in order to learn a naive Bayes classifier, one begins by assuming that values conditional on class are probabilistically independent. This assumption is represented in the "Bayesian network" shown in Figure 2 (a). Similar to the DAGs introduced above in the context of causal models, a Bayesian network is a compact representation of statistical independence relations. Random variables are represented as nodes, and edges are drawn such that each node is independent of all its non-descendants conditional on its parents. Violations of the naive Bayes assumption suggest the existence of latent variables left out of one's initial DAG. If we assume that the true probabilistic relationships amongst the observed variables and the as yet unknown latents is hierarchical (i.e., the DAG is a tree as shown in Figure 2 (b) then it is possible to learn the existence of the latent variables by searching through the space of all possible hierarchical Bayesian models compatible with the observed class and feature variables. The search begins with the simplest model, i.e., that learned by the naive Bayes algorithm. At each stage of the search, new candidate models are generated by adding new nodes intermediate to existing nodes and their children, shifting the parents of existing nodes, and deleting nodes. Each new model is assessed using the Bayesian information criterion (Schwarz, 1978), a measure that offsets the posterior probability the model assigns to the data by a penalty that depends on the number of free parameters in the model. The optimal model found by this algorithm, assuming that in fact the naive Bayes assumption is false, will contain latent variables. In other words, the algorithm posits the existence of unobserved variables. That looks a lot like the sort of conceptual novelty demanded by the deep structure objection. In fact, many algorithms now routinely deployed in ML posit latent variables to account for variation or patterns in given data.[15]

The second algorithm I wish to highlight as a prima facie rebuttal to the deep structure objection is from the BACON lineage (Langley et al, 1987), specifically BACON.4. This algorithm, when confronted with properties that vary across individual systems (a fact encoded in the data by introducing a discrete nominative variable to label each system) posits a new *intrinsic property* that varies from individual to individual. This unobserved property is not directly assessable by observation, and is not formulated in terms of the given variables – it is as much a 'novel concept' in Hempel's sense as one could hope for.

To see how this works, consider the data shown in Table 1. It represents measurements made on wires connected to various batteries. Letters label distinct wires and batteries. The only variable measured is

---

[15] Two of the most prominent are the mixture of Gaussians and Independent Component Analysis (ICA) algorithms.
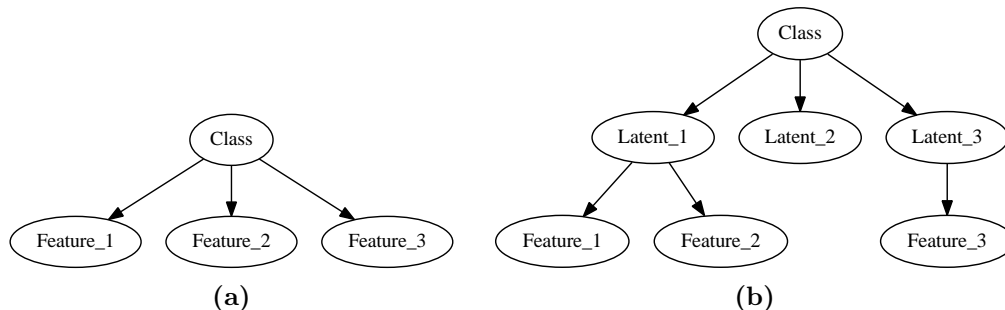
Fig. 2 (a) A representation of a naive Bayes model as a directed graph. (b) A hierarchical naive Bayes (HNB) model with a single layer of latent variables.

Battery	Wire	Current (i)	Conductance (c)	Voltage (c)
A	X	3	3	1
A	Y	4	4	1
A	Z	5	5	1
B	X	6	3	2
B	Y	8	4	2
B	Z	10	5	2
C	X	9	3	3
C	Y	12	4	3
C	Z	15	5	3

Table 1 Data used by the BACON.4 algorithm to introduce the intrinsic variables of conductance and battery voltage. This is a modification of an example given in (Langley et al, 1987).

current, which is generally represented by a real number, though I've simplified the example to use all integer values. Upon examining the first three rows of the table (corresponding to the same battery but different wires), BACON notes that current varies from wire to wire. The next step of the algorithm is, practically speaking, driven by the fact that BACON cannot relate non-numerical variables (e.g., the identifiers for distinct wires) to numerical variables. But we might give it a rather plausible methodological interpretation: if a variable changes from one circumstance to the next – in this case, from one wire to the next – it is reasonable to suppose that there exists a hidden, causally salient property which varies from wire to wire. Let's call that property 'conductance', and assume that it can be represented by a real number as well.

Following this maxim, BACON introduces a *new variable* whose values are shown in the third column. How were these values determined? As is clear from the table, BACON assigns a conductance equal to the values of the previously known variable, current. The authors don't discuss this procedure much, but it is a simple way to ensure that the new variable explains the old in the sense that there is a unique conductance value for each resulting current.

So far, it's not clear that the "new" variable is very informative or novel. But things get interesting when we get to the next three rows of the table. Since each wire was already assigned a value for conductance, BACON uses those values again, and notes that for battery B, the conductance and the current are proportional to one another. Unlike the case for battery A, however, the constant of proportionality is now 2. Similarly, for the last three rows (corresponding to battery C), BACON finds that conductance and current are related by a slope of 3. How to explain this variation? Posit a new variable! This time, we suppose there is a property of batteries (the voltage) that explains the variation, and we assign values identical to the slopes in question. If we note that conductance is the reciprocal of resistance, we can see that BACON has just 'discovered' Ohm's law of resistance (r): $i = vc = v/r$. Of course, that relation is tautological if we consider only the data on hand. But treated as a generalization, it is quite powerful and most definitely falsifiable. We might, for instance, find that a new wire, D, has a conductance of c as determined using battery B. But when connected to battery A, the new wire could show a current not equal in value to c. This would violate Ohm's law.

With BACON.4 as with the preceding algorithm, we have a method for positing law-like relations in terms of novel variables. Why hasn't this been decisive? Why haven't philosophers of science turned their attention in any significant way to logics of discovery? Perhaps the results of ML have simply failed to penetrate the philosophical literature. Perhaps it is because ML models don't look like the kinds of theories philosophers of science are used to engaging. Or, less superficially, perhaps there is still the sense that learning algorithms are dispensable. After all, the thinking goes, science did quite well before ML. Sure it's great if we can more efficiently feed theories into the mill of post-hoc theory testing, but it isn't necessary. The logic of justification can take care of everything. Eventually, generate and test will get us to the truth. In the remainder of this essay, I will demonstrate that this line of thinking is untenable. In fact, it is the very difficulty of the task of identifying novel variables – the task which foes of discovery fingered as beyond algorithmic reach – that entails the need for algorithmic approaches. In other words, success at deep structure discovery demands the application of a substantive logic of discovery.

4 The problem of 'deep structure' discovery

4.1 What's the problem?

There are many methodological contexts that provoke the introduction of novel variables into scientific theory and practice, and many roles such variables play. To make my case for a logic of discovery, I'll focus on just one aspect of the challenge of identifying deep structure raised above. In particular, I want to restrict attention to the generation of hypotheses that contain new variables where those hypotheses are intended to cover an established domain of investigation. In this case, one has already characterized the phenomena to be explained in terms of some set of preexisting variables that have themselves proved insufficient to the task of explanation or prediction in one way or another. So, for instance, naked-eye observational astronomy offers a host of phenomena to be explained, including the motion of the sun and planets against the background stars, the periodic retrograde motion of the planets, the different periods with which they execute their angular motions, the curious relation between solar years, cycles of anomaly, and revolutions of longitude for the superior planets, and many other such facts, all couched in terms of angular positions, time, and brightness that together comprise an initial set of variables, V. The problem of 'deep structure' identification in such cases is to posit new variables that explain these phenomena, given one's preferred notion of explanation, and which make accurate predictions about both old and new variables. In other words, we posit a new set of variables, V', and law-like relations amongst the variables in V' and those in V such that the new variables V' can be used to explain and predict aspects of the phenomena of interest, now described in terms of the full set $V \cup V'$. This is what Ptolemy, Copernicus, and Kepler did when they introduced their respective theoretical variables. These include periods, sizes, and relative distances between the centers of cycles, epicycles, and equants for Ptolemy, a different set of theoretical cycle and epicycle periods, sizes, and distances for Copernicus, and a whole new set of periods and orbital parameters relative to the sun for Kepler. Each of their respective hypotheses made use of new variables in order to explain patterns in celestial motion, as well as to accurately predict that motion.

To give another example, consider Pauli's postulation of the neutrino. Beta decay — a process in which a nucleus emits an electron and converts a neutron to a proton — was a well characterized phenomenon, described in terms of nucleon number, mass, charge, angular momentum, linear momentum, rest energy, and kinetic energy. What was lacking was an explanation for why the electrons emitted in the process exhibit a continuous range of kinetic energies, a fact which seemed to violate the conservation of energy. Pauli proposed the existence of a new particle type — the neutrino — and thus introduced a collection of new variables, namely the rest mass, momentum, and kinetic energy of this new particle. The energy of the emitted neutrino explained the continuous energy spectrum of emitted electrons, and predicted various phenomena as consequences of the properties of the postulated particle. This is a paradigm instance of discovering or positing what Laudan calls deep structure.

The problem of identifying deep structure is thus the problem of choosing a suitable collection of explanatory variables and, simultaneously, a set of laws linking new and old variables that makes true predictions. Before considering the difficulty of this problem, I will here again make my task easier by simplifying the task of the scientist: I will focus entirely on predictive accuracy. That is, I will consider only the problem

of generating an hypothesis that references new kinds of variables and which is predictively successful over some significant range of new data. Such an hypothesis may or may not be explanatory in one's preferred sense, and so may fail to really get at the 'deep structure'. But again, adding the additional constraint of being explanatory only makes the problem of generation harder (assuming that explanatory hypotheses are *also* predictive). If I can make my case for the less restrictive problem of identifying merely predictive deep structure, then my argument is only strengthened by considering the full problem.

4.2 How difficult is the problem?

Given that we are only worried about predictive success, how much of a problem does this really pose? Since there is no definite limit on the number or nature of variables one might introduce; the possibilities are infinite. Pauli might have suggested two kinds of particle, or three. Or he might have posited a new field akin to the electromagnetic field that carries away varying portions of the decay energy. Or he might have suggested some entirely novel variables. Infinitely many of the possible choices of variable would have been useless, irrelevant, and — if one has realist inclinations — not even approximately true. Conversely, there were infinitely many choices Pauli might have made that would have worked just as well, at least in the short run (this is run of the mill underdetermination). By itself, these infinities don't tell us much about the difficulty of selecting a successful hypothesis. What matters is the relative 'size' of the sets of good and bad hypotheses.

The space of all possible hypotheses is ill-defined, and cleaning it up would require choosing a clear position with respect to some difficult problems, like whether or not to be a realist about theoretical terms and what the general relationship is between theoretical variables and observational variables in any acceptable theory. Nonetheless, there are some general considerations that suggest that the problem of novel variable choice is very hard. Both the space of predictively successful and the space of unsuccessful hypotheses are presumably uncountably infinite, and thus of the same cardinality. The relevant notion of size is therefore not cardinality but something like a measure in the mathematical sense.[16] Such a measure gives us a notion of volume that is appropriate for, say assessing relative frequency or probability. To determine the relative sizes of the sets of good and bad hypotheses in this way, consider instead sequences of predicted outcomes. That is, consider some finite sequence of predictions made by an hypothesis. In the simplest case, consider just one prediction. For the hypothesis to be successful, this prediction must be accurate. To simplify things further, suppose that each prediction involves specifying the value of a single continuous variable (i.e., a variable whose possible values are well represented by the real numbers). For instance, we might consider predictions of the mass of the Higgs boson according to each published hypothesis and every logically possible alternative.[17] To be realistic, we can further suppose that the 'correct' answer is anything within some open interval.[18] That is, there is some range of values the hypothesis might predict such that any value in that range is 'close enough'. Now, the space of all possible values is the union of an infinite number of such intervals. If our measure – our volume function – treats every interval equally, then the set of correct answers is vanishingly small relative to the set of wrong answers. That is, the set of right answers receives some finite value for its size, while both the set of all answers and the subset of wrong answers receive infinite values. What does this tell us about hypotheses? If we assume that for every possible interval of prediction values there is an equal proportion of possible hypotheses that predict a value in that interval, then the set of successful hypotheses is vanishingly small.[19] In fact, the proportion of successful hypotheses is 0.

---

[16] Let S be a set. A σ-algebra is a non-empty set Σ such that: (1) S is in Σ, (2) if A is in Σ then so is the complement of A, and (3) the union of any sequence of elements of Σ is in Σ. A measure is a function from Σ to the positive real numbers such that the null set maps to zero and such that the measure of the union of any finite or countably infinite set of disjoint members of Σ is equal to the sum of the measures of each set alone

[17] In a minority of cases, the definition of a variable (*not* substantive knowledge of its possible values) ensures that the space of logically possible values it can assume is bounded. For example, celestial longitude can only take on values between 0 and 360 degrees. In those cases, the argument of this section does not go through; there will be some finite, nonzero probability of guessing a new value correctly. I stress, however, that most physical variables take on an unbounded range of possible values.

[18] For technical reasons, it's easier to consider left closed, right open intervals. The claims made in the text presume intervals of this sort.

[19] For various sorts of anti-realist, this is a plausible assumption, since hypotheses are individuated by their empirical consequences. Thus, each prediction counts for one hypothesis. The realist, however, individuates hypotheses based on what they say concerning observable and unobservable entities. In that case, I have perhaps made a dubious assumption. But no more

In its barest terms, even the weak problem of identifying predictive deep structure is breathtakingly hard in the sense that it's not something one can succeed at accidentally. From an infinity of possibilities, we need to identify a hypothesis from the vanishingly small proportion that successfully predict. If this problem seems trivial, it is only because we are used to taking its unspoken solution for granted.

5 Why the actual practice of science must rely on a logic of discovery

5.1 Hypothesis generation cannot be random

As I suggested at the outset, the difficulty of the problem of identifying novel variables bears on the question of whether there is a non-trivial logic of discovery of relevance to *actual* scientific practice. To make the connection, suppose there is no logic of discovery as I've defined it. What would this entail? There are a number of ways one might understand the denial of a logic of discovery. In the strongest and least plausible view, it amounts to the claim that the production of hypotheses is stochastic. As Shah succinctly put it, "The opposite of a logic of discovery is a process of hypothesis generation that is completely random"(Shah, 2008, p308). By "completely random", I (and, I think, Shah) mean that the process of theory generation involves an arbitrary draw from the entire space of logical possibilities. This is an extreme position that looks a lot like a straw man, and I'll set to work modifying it in short order. But it bears attention for two reasons. First, it often appears that the foes of discovery do in fact have something like this in mind. Consider how Curd (1980, pp 206-7) states the matter:

> Now at first blush, the denial of a logic of theory generation seems a counterintuitive conclusion. Are scientists just guessing when they come up with hypotheses? One has visions of that mythical tribe of chimpanzees which, by bashing away for long enough on a stack of typewriters, would eventually generate not only the complete works of Shakespeare but also Einstein's general theory of relativity and the Watson-Crick hypothesis for the structure of DNA. The position that the hypothetico-deductive account adopts with respect to this issue is not the result of prejudice but is based on arguments which many philosophers of science still find convincing.

The arguments against the possibility of a logic of discovery to which he alludes were dealt with in Section 2. My intended emphasis in the quoted passage is rather on Curd's characterization of the alternative to a logic. It suggests that at least some of those who deny the possibility of a logic of discovery really do envisage theory generation as something close to a random draw, as astonishing as that may seem. The reason it seems astonishing is straightforward. If hypotheses are generated at random from the space of possibilities open at any one time, then the probability of generating a successful hypothesis is vanishingly small, at least in the case of identifying Laudan's 'deep structure'. As I noted above, the space of possible hypotheses (possible postulations of explanatory variables and their relations to existing variables) is infinite for any such problem, and any plausible measure over this space would attribute a relative measure of zero to the subset of suitable hypotheses. In other words, the probability (coarsely speaking) of randomly choosing a successful hypothesis containing novel variables is nil.[20] Yet individual scientists and the scientific community do this all the time. Thus, if the foes of discovery are right, then the success of science is nothing short of a miracle. If one rejects miraculous explanations as I do, then the success of science implies the existence of a logic of discovery.

5.2 Even constrained random generation requires a logic

Barring miracles, it cannot be the case that hypotheses are generated at random. A tempting rejoinder to this line of reasoning is to point out that hypotheses are drawn from only a subset of all the logical possibilities. The vast space of possible variable identifications is irrelevant, one might suppose, because the scientist is never dealing with its entirety. Though hypotheses may be produced via psychological processes that are more or less random, we only ever need to sample from a small space of hypotheses. Why? Because preexisting

---

can be said unless we know something about how to define the set of possible hypotheses and what to treat as equal regions of hypotheses.

[20] There is no probability measure obeying the Kolmogorov axioms of probability that is uniform over the entire real line. But one can assign a measure over the real line and ask about the relative size of subsets of the reals with respect to that measure.

ontological commitments, theoretical virtues like simplicity or fecundity, and consistency with other theories all serve to narrow the possibilities. So even without an algorithmic method for producing hypotheses, there would be no miracle. Surely, in such a constrained space of possibilities, a simple generate-and-test approach will succeed in short order, no matter how random the generation step.

Under scrutiny, however, this view fares no better. To begin with, the history of science suggests that scientists are quite good at generating predictively successful hypotheses. As Thagard (1980) has pointed out in an argument against evolutionary epistemology, the history of science shows that the rate at which successful hypotheses are produced relative to all hypotheses is strikingly high. In other words, scientists are not producing large numbers of hypotheses that need to be winnowed to find a good one. This fact implies that however hypotheses are being produced, the pared-down sets of candidate hypotheses must contain a rather high ratio of short-run projectible to unprojectible hypotheses in order to account for the observed rates of success. In other words, the pool of hypotheses from which scientists are drawing at random must consist mostly of hypotheses that are successful in the sense of being short-run projectible. Of course, terms like "mostly" and "high" are vague. But nothing much rests on where exactly one draws the line. The point is that if we are to account for the rate of success in generating deep structure hypotheses, we must concede that the set of candidate hypotheses from which we sample at random (if indeed we sample at random) must be very rich in short-run projectible hypotheses.

But the degree to which the space of possible hypotheses must be whittled down before one has a set that is sufficiently enriched with successful hypotheses to make a random generate-and-test approach work in a reasonable amount of time is extreme. The pared-down set of candidates from which we draw must be of finite measure, or else the ratio of successful to total hypotheses even in the pared-down set would simply be zero. But that means that in order to pare down the space of logical possibilities in order to get a set for which guessing is efficient requires selecting a subset of the possibilities of relative measure zero in the first place. Picking the right subspace to guess from without a method is just as miraculous as picking a projectible hypothesis when done at random.

But surely this too is unfair. Scientists don't choose a set of candidate hypotheses (consciously or not) at random. The point is that the pool itself is dictated both by normative and practical constraints. On the practical side, we do not consider hypotheses that would take centuries to write down, or which we cannot assess for consistency. On the normative side, appeals to existing ontological commitments, consistency with well-established theory, and theoretical virtues such as simplicity all serve to narrow the space of candidate hypotheses. Of course, if these constraints are to do any work in actually generating a set of candidate hypotheses (from which random selection and testing is good enough), they must be generally applicable. In other words, they must take the form of domain-general rules to be of any use. If the constraints vary from application to application, then the problem of identifying the constraints looks as hard as the problem of identifying variables. If, for instance, theoretical virtues like simplicity are to be capable of narrowing the space of possible hypotheses in the relevant way, it must be a decidable question whether an arbitrarily chosen hypothesis is sufficiently simple, regardless of whether that hypothesis is biological, chemical, or physical in nature.

But by now, the proposed alternative to a logic of discovery just *is* a logic of discovery. Shah (2008) makes this point about Popper's later view of the process of science. Roughly, Popper suggests that science proceeds by a series of conjectures and refutations that resembles biological evolution. In both science and evolution, the "method of trial and error-elimination *does not operate with completely chance-like or random trials* (as has been sometimes suggested), even though the trials may look pretty random; there must be at least an 'after-effect'"(Popper, 1972, p245, fn55)(emphasis in the original). A sequence of events exhibits an "after-effect" in Popper's sense if the probability of a particular outcome at position n in the sequence is not independent of some finite segment of preceding events. In this case, it is an acknowledgement that the generation of new hypotheses must depend upon what has been tried and refuted before—there must be a constraint on the space of possibilities. Popper is circumspect about what these constraints may be, but asserts elsewhere that

> ...to understand a problem means to understand its difficulties; and to understand its difficulties means to understand why it is not easily soluble—why the more obvious solutions do not work. We must therefore produce these more obvious solutions; and we must criticize them, in order to find out *why* they do not work (Popper, 1972, p260).

The idea seems to be that ways in which a hypothesis fails may be used to constrain the space of plausible substitutes. One could imagine, for instance, adjusting theories in the face of empirical conflict using a list of rules, e.g., "given a conflict with experiment that can be traced to a particular functional relation, adjust that relation by fitting the simplest function that encompasses the new data." Whatever the specific methods Popper might have had in mind, the point is that they constitute a logic of discovery, an algorithm that takes us from one failed hypothesis to a *very* restricted class of good alternatives from which a random draw is likely to yield success. While Shah argues that Popper's prescribed methodology in particular should be seen as a logic of discovery, my point is that *any* set of constraints general enough and strong enough to present a set of alternatives that is rich enough in good hypotheses to make random guessing succeed just is a logic of discovery.

5.3 Is the question of a method all or nothing?

At this point, one might worry that I am still trading on a false dichotomy. Surely there is a lot of daylight between a completely deterministic *algorithm* for generating hypotheses, and a totally random guess. While the use of constraints or heuristics are not purely random, nor are they the sort of thing we could write an algorithm for. At least that's what authors such as Kleiner (1993) have suggested. But is it true? To answer the question, we need to be more precise about the supposed alternatives. One can represent these supposed intermediate possibilities by, for instance, considering deterministic algorithms that map past observations, etc. onto probability distributions over the space of possible hypotheses. Those distributions may be more or less concentrated around a particular hypothesis. At one end of the spectrum are algorithms–step-by-step procedures–that map each sequence of observations to probability distributions concentrated on a single hypothesis (or infinitesimal neighborhood); these are in some sense maximally deterministic. At the other end of the spectrum are cases in which hypothesis generation is essentially random in that the algorithmic step yields only an extremely diffuse distribution over a large space of hypotheses (as close to uniform as possible). By denying a logic of discovery, you might suggest that one is only denying the deterministic extreme. Perhaps, it is conceded, there is some weak role for algorithmic discovery procedures, but the important steps are random. However, these intermediate possibilities are illusory. As we've already seen, the important steps are those which pare down the initial space of possibilities, or in other words, the step that moves from past observations to a probabilistic distribution spread over a modest subset of these possibilities. Furthermore, there is nothing to the random steps – whether important or not – that can't be achieved with a deterministic algorithm. Suppose we can account for success at discovery by appealing to a combination of a deterministic function from observations to a distribution over hypotheses according to which we make random draws. Then there exists a completely deterministic function that does just as well. That's because for any random draw, we could use the rejection sampling algorithm along with a pseudorandom number generator to select one particular hypothesis. This new procedure is both deterministic and overtly algorithmic, since both the requisite pseudorandom generator and the rejection sampling algorithm can be accomplished with known deterministic algorithms. Furthermore, this deterministic algorithm must do as well as the genuinely random draw, since it is indistinguishable in its output across all possible inputs. Thus, if a supposedly intermediate option exists, so too does an equally good, deterministic option. If one wants to deny the existence of a logic of discovery by denying the existence of an algorithmic method, one has to embrace fully random guessing and the miracles that entails.

5.4 Content-specific heuristics, but no content-neutral logic?

So far, I have been arguing as though the choice is between a single overarching method and no method at all. This is at odds with a view popular since the early twentieth century that denies the existence of any algorithmic, domain- or content- neutral logic of discovery, while recognizing myriad domain- or content-specific heuristics that do the job of regimenting discovery within particular sciences (see, e.g., Carmichael, 1922; Nickles, 1990). How exactly are content-specific heuristics supposed to explain scientists' success at the task of identifying deep structure? The idea is that there are specific strategies for positing new variables in, say, high-energy physics, and different, content-rich strategies in genomics, and so on. In each context, there is a set of rules or procedures that are "suggestive rather than demonstrative" (Schickore, 2014) and which

refer essentially to theoretical or empirical facts about that context. Most suggested heuristics of this sort comprise stereotyped versions of reasoning processes that have yielded success in the past. To give a concrete example, there is the strategy in theoretical physics of constructing interacting quantum field theories by promoting a known or hypothesized global symmetry to a local gauge symmetry. This strategy will not always work, but it's worked enough times to be worth trying in new cases.

Each heuristic amounts to a strong restriction on possible hypotheses, where the particular set of candidates is determined by content, i.e., past theories in a particular domain, previous predictions, empirical facts about the relevant domain, and so forth. As I argued above, these heuristics must be quite strong if they are to explain success at identifying novel variables. Is it plausible that such heuristics exist without any global, content-neutral logic? As with the constraints considered above, the choice of a heuristic is the selection of a subset of all possible hypotheses with a relative size of zero. So how are these choices made? Where do the heuristics come from? If they are selected at random for each empirical domain, it would be just as much a miracle that they work as it would be to randomly select a successful hypothesis. Thus, we must assume that the choice of heuristic is governed by a meta-heuristic. The meta-heuristics must of necessity be rather more content-neutral than the heuristics they help us to select. But if it's heuristics all the way down, with each layer of meta-heuristic increasingly content-neutral, then we've simply described multiple stages of a general, iterative logic of discovery. More abstractly, if we are guided at every step of the way in developing our heuristics (so as to avoid a miracle), then we must have a procedure that begins in a quite content-neutral fashion. As one iteratively applies the procedure, each step becomes increasingly content-specific. Whatever the exact form of the overarching logic, the upshot is that admitting ubiquitous heuristics to explain success at the deep structure problem is tantamount to admitting a general, domain-neutral logic of discovery. Otherwise, we have only replaced one sort of miracle with another.

5.5 Logic and computability

Finally, there is what I take to be a far more charitable interpretation of the claim that no methodologically relevant logic of discovery exists. To draw an important distinction, let's call any function that (i) is implementable by human beings, and (ii) maps observations, past successes, theoretical commitments, etc. to one of the subsets of logical possibilities that is rich in successful hypotheses a *discovery function*. Put differently, a discovery function is any deterministic method that meets all of the conditions I gave for a logic of discovery with the possible exception of computability. We can now see that it's the latter condition that invites debate. Rather than objecting to the existence of a deterministic method, the foe of discovery may simply intend to deny that the relevant discovery function is computable. While our brains might be able to leap to the output of the relevant function (and in that sense, 'implement' the function), we cannot represent this function via a mechanical procedure. In that case, there is nothing worth calling a logic,[21] and certainly nothing we could use to automate the process of theory generation. However, if that's the thesis of the foe of discovery – that although there is indeed a deterministic discovery function we (approximately) implement to move from data to successful hypotheses, it is not computable – then the debate over whether or not a logic of discovery exists collapses onto the debate over whether humans (or communities of humans) are capable of completing uncomputable tasks. If you think we cannot, then the implemented function must be computable, and there must be a logic of discovery for the reasons given above. The only way to consistently deny the existence of such a logic is to accept that humans can do the uncomputable. There are, of course, plenty who would argue for this claim (see, e.g., Penrose, 1994). However, there seem to be equally cogent reasons for denying it (see, e.g., Feferman, 1995). The point I want to stress is that the locus of debate is not where the foes of discovery typically place it. The real question is whether or not something like the computational theory of mind is correct.

---

[21] I do not intend to concede this point. As Penrose (1994) points out, we can represent and reason about uncomputable functions. To say there is nothing worth calling a logic of discovery given an uncomputable function that otherwise has all of the features of such a logic requires an argument. Presumably, such an argument would turn on the claim that to count as a logic of discovery, it must be possible to consciously implement the procedure, and that this can only be done for computable functions. Whatever we do that's uncomputable, it is not and cannot be conscious, hence the talk about happy guesses and strokes of genius.

6 Conclusion

My argument for the existence of at least one methodologically relevant logic of discovery is straightforward. A common scientific problem is to identify so-called 'deep structure' – novel variables and laws connecting them with old variables in such a way that one is able to accurately predict what could not be predicted (or explained) before. There are two mutually exclusive and exhaustive possibilities: either (i) there exists a discovery function (not necessarily computable) that is implementable by human beings and that maps observations, past theoretical posits, etc. to a subset of all logically possible hypotheses such that each such subset contains a non-vanishing proportion of hypotheses with limited projectibility; or (ii) there does not exist such a function. I have argued that this poses a dilemma. If there does not exist a discovery function, the success of science is miraculous in that there is a probability (in the sense of a relative measure) of zero of choosing a hypothesis that correctly projects the value of a single real-valued variable at random. If there is a discovery function, scientific success is not miraculous, but we must accept the existence of a method against which philosophers of the twentieth century spent a great deal of time arguing.

I take it as obvious which horn of the dilemma we should embrace, given the general implausibility of miracles. But if one were so inclined, this dilemma could be filled out as an inference to the best explanation. In other words, we have a fact to be explained, namely the success of science. There are two hypotheses available: either there is or is not a discovery function. The hypothesis that there is not such a function makes the fact to be explained virtually impossible, whereas the hypothesis that there exists a discovery function grants the success of science a vague (unless we specify the function) but significant probability. Therefore, we should (overwhelmingly) favor the hypothesis that there exists a discovery function which scientists follow, at least approximately, in order to succeed at deep structure discovery.

Note that this inference to the best explanation is immune to Van Fraassen's (1990) "best of a bad lot" objection because the alternatives under consideration are exhaustive. Furthermore, the 'probability' (relative measure) of the explanandum given one of the hypotheses is literally zero. The conclusion that a discovery function exists is thus as close to morally certain as one can get. Note as well that the conclusion is not merely historical. As far as scientists were successful at deep structure identification in the past, and as far as they wish to be successful at it in the future, they must approximate a discovery function in their actions. There *must* exist a discovery function that is approximated by actual scientific practice, past and future. So unless one is willing to concede that scientists achieve the uncomputable, the routine success of scientists at identifying deep structure implies – with practical certainty – the existence of a logic of discovery.

Of course, one philosopher's modus ponens is another's modus tollens. Even if you embrace the existence of a discovery function, you might still deny the existence of a logic of discovery as I defined it, namely a discovery function that is computable. In particular, you might take this argument as an indictment of the computational theory of mind – there must exist a discovery function but it is implausible that such a thing is computable. If that's the case, then humans must (individually or collectively) accomplish the uncomputable. Determining the plausibility of that position is beyond the scope of this paper. So too is attempting an explicit characterization of a logic of discovery. Of course, there may be more than one. Not in the sense of numerous, domain specific heuristics, but rather in the sense of multiple global strategies for producing hypotheses with limited projectibility. A good place look for an explicit formulation of such a thing is the current machine learning literature. Whether the seeds of an answer lie there or elsewhere, we should be looking. Given that we have good reason to believe that a logic of discovery exists and is essential to scientific practice, articulating that method is thus of paramount importance. The logic of discovery belongs front and center in the philosophy of science.

References

Battle A, Jonikas MC, Walter P, Weissman JS, Koller D (2010) Automated identification of pathways from quantitative genetic interaction data. Molecular Systems Biology 6(1):n/a–n/a, DOI 10.1038/msb.2010.27, URL http://onlinelibrary.wiley.com/doi/10.1038/msb.2010.27/abstract

Bengio Y (2009) Learning Deep Architectures for AI. Foundations and Trends in Machine Learning 2(1):1–127, DOI 10.1561/2200000006

Bishop C (2007) Pattern Recognition and Machine Learning. Springer, New York

Breiman L, Friedman J, Stone CJ, Olshen RA (1984) Classification and Regression Trees, 1st edn. Chapman and Hall/CRC, Boca Raton

Carmichael RD (1922) The Logic of Discovery. The Monist 32(4):569–608, URL http://www.jstor.org/stable/27900927

Caschera F, Gazzola G, Bedau MA, Bosch Moreno C, Buchanan A, Cawse J, Packard N, Hanczyc MM (2010) Automated Discovery of Novel Drug Formulations Using Predictive Iterated High Throughput Experimentation. PLoS ONE 5(1):e8546, DOI 10.1371/journal.pone.0008546, URL http://dx.doi.org/10.1371/journal.pone.0008546

Craver CF, Darden L (2013) In Search of Mechanisms: Discoveries across the Life Sciences. University Of Chicago Press, Chicago ; London

Curd MV (1980) The logic of discovery: An analysis of three approaches. In: Cohen RS, Wartofsky MW, Nickles T (eds) Scientific discovery, logic, and rationality, Boston Studies in the Philosophy of Science, vol 56, D. Reidel Publishing Company, Dordrecht, Holland, pp 201–219

Darden L (2002) Strategies for Discovering Mechanisms: Schema Instantiation, Modular Subassembly, Forward/Backward Chaining. Philosophy of Science 69(S3):S354–365, DOI 10.1086/341858, URL http://www.jstor.org/stable/10.1086/341858

Darnell SJ, Page D, Mitchell JC (2007) An automated decision-tree approach to predicting protein interaction hot spots. Proteins: Structure, Function, and Bioinformatics 68(4):813–823, DOI 10.1002/prot.21474, URL http://onlinelibrary.wiley.com/doi/10.1002/prot.21474/abstract

Ebert-Uphoff I, Deng Y (2012) Causal Discovery for Climate Research Using Graphical Models. Journal of Climate 25(17):5648–5665, URL http://search.proquest.com/docview/1039516026

Feferman S (1995) Penroses Gdelian argument. Psyche 2(7)

Feigenbaum EA, Buchanan BG (1993) DENDRAL and Meta-DENDRAL: roots of knowledge systems and expert system applications. Artificial Intelligence 59(12):233–240, DOI 10.1016/0004-3702(93)90191-D, URL http://www.sciencedirect.com/science/article/pii/000437029390191D

Flach P (2012) Machine Learning: The Art and Science of Algorithms That Make Sense of Data. Cambridge University Press, URL http://proquest.safaribooksonline.com/9781107096394

van Fraassen BC (1990) Laws and Symmetry. Oxford University Press, Oxford ; New York

Freno A (2009) Statistical Machine Learning and the Logic of Scientific Discovery. Iris: European Journal of Philosophy & Public Debate 1(2):375–388, URL http://ezproxy.lib.vt.edu:8080/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=hlh&AN=47365071&scope=site

Gillies D (1996) Artificial intelligence and scientific method. Oxford University Press, Oxford [England] ; New York

Gutting G (1980) Science as Discovery. Revue Internationale de Philosophie 131:26–48

Han LY, Zheng CJ, Xie B, Jia J, Ma XH, Zhu F, Lin HH, Chen X, Chen YZ (2007) Support vector machines approach for predicting druggable proteins: recent progress in its exploration and investigation of its usefulness. Drug Discovery Today 12(78):304–313, DOI 10.1016/j.drudis.2007.02.015, URL http://www.sciencedirect.com/science/article/pii/S1359644607000967

Hanson NR (1961) Patterns of discovery: an inquiry into the conceptual foundations of science. University Press, Cambridge [Eng.]

Hempel CG (1966) Philosophy of natural science. Prentice-Hall foundations of philosophy series, Prentice-Hall, Englewood Cliffs, N.J

Hoyningen-Huene P (2006) Context of discovery versus context of justification and Thomas Kuhn. In: Schickore J, Friedrich Steinle (eds) Revisiting Discovery and Justification, no. 14 in Archimedes, Springer Netherlands, pp 119–131, URL http://link.springer.com.ezproxy.lib.vt.edu/chapter/10.1007/1-4020-4251-5_8

Kelly KT (1987) The Logic of Discovery. Philosophy of Science 54(3):435–452, URL http://www.jstor.org/stable/187583

Kenny LC, Dunn WB, Ellis DI, Myers J, Baker PN, Kell DB (2005) Novel biomarkers for pre-eclampsia detected using metabolomics and machine learning. Metabolomics 1(3):227–234, DOI 10.1007/s11306-005-0003-1, URL http://link.springer.com/article/10.1007/s11306-005-0003-1

King RD, Whelan KE, Jones FM, Reiser PGK, Bryant CH, Muggleton SH, Kell DB, Oliver SG (2004) Functional genomic hypothesis generation and experimentation by a robot scientist. Nature 427(6971):247–252, DOI 10.1038/nature02236, URL http://www.nature.com/nature/journal/v427/

n6971/full/nature02236.html

Kleiner SA (1993) The logic of discovery: a theory of the rationality of scientific research. Kluwer Academic Publishers, Dordrecht; Boston

Langley P, Zytkow JM (1989) Data-driven approaches to empirical discovery. Artificial Intelligence 40:283–312, DOI 10.1016/0004-3702(89)90051-9, URL http://www.sciencedirect.com/science/article/pii/0004370289900519, 13

Langley P, Simon HA, Bradshaw G, Zytkow JM (1987) Scientific discovery: computational explorations of the creative processes. MIT Press, Cambridge, Mass

Laudan L (1980) Why was the logic of discovery abandoned? In: Nickles T (ed) Scientific discovery, logic, and rationality, Boston Studies in the Philosophy of Science, vol 56, D. Reidel Publishing Company, Dordrecht, Holland, pp 173–183

Laudan L (1981) Science and Hypothesis. D. Reidel Publishing Company, Dordrecht, Holland

Maron ME, Kuhns JL (1960) On Relevance, Probabilistic Indexing and Information Retrieval. J ACM 7(3):216–244, DOI 10.1145/321033.321035, URL http://doi.acm.org/10.1145/321033.321035

McLaughlin R (1982) Invention and Induction Laudan, Simon and the Logic of Discovery. Philosophy of Science 49:198–211, DOI 10.2307/186918, URL http://www.jstor.org/stable/186918, 2

Medina F, Aguila S, Baratto MC, Martorana A, Basosi R, Alderete JB, Vazquez-Duhalt R (2013) Prediction model based on decision tree analysis for laccase mediators. Enzyme and Microbial Technology 52(1):68–76, DOI 10.1016/j.enzmictec.2012.10.009, URL http://www.sciencedirect.com/science/article/pii/S0141022912001779

Michalski R, Chilausky R (1980) Knowledge acquisition by encoding expert rules versus computer induction from examples: a case study involving soybean pathology. International Journal of Man-Machine Studies 12(1):63–87, DOI 10.1016/S0020-7373(80)80054-X, URL http://www.sciencedirect.com/science/article/pii/S002073738080054X

Mjolsness E, DeCoste D (2001) Machine Learning for Science: State of the Art and Future Prospects. Science 293(5537):2051–2055, DOI 10.1126/science.293.5537.2051, URL http://www.sciencemag.org/content/293/5537/2051

Nickles T (1984) Positive Science and Discoverability. PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association 1984:13–27, URL http://www.jstor.org/stable/192324

Nickles T (1985) Beyond Divorce: Current Status of the Discovery Debate. Philosophy of Science 52(2):177–206, URL http://www.jstor.org/stable/187506

Nickles T (1990) Discovery Logics. Philosophica 45:7–32

Pearl J (2000) Causality : models, reasoning, and inference. Cambridge University Press, Cambridge, U.K.; New York

Penrose R (1994) Shadows of the mind: a search for the missing science of consciousness. Oxford University Press, Oxford ; New York

Popper K (1972) Objective Knowledge. Oxford University Press, London

Popper K (2002) The logic of scientific discovery. Routledge Classics, Routledge, New York

Quinlan JR (1992) C4.5: Programs for Machine Learning, 1st edn. Morgan Kaufmann, San Mateo, Calif

Reichenbach H (1938) Experience and prediction: An analysis of the foundations and the structure of knowledge, vol x. University of Chicago Press, Chicago, IL, US

Russell S, Norvig P (2009) Artificial Intelligence: A Modern Approach, 3rd edn. Prentice Hall, Upper Saddle River

Savary C (1995) Discovery and Its Logic: Popper and the "Friends of Discovery". Philosophy of the Social Sciences 25(3):318–344, DOI 10.1177/004839319502500303, URL http://pos.sagepub.com/content/25/3/318

Schickore J (2014) Scientific Discovery. In: Zalta EN (ed) The Stanford Encyclopedia of Philosophy, spring 2014 edn, URL http://plato.stanford.edu/archives/spr2014/entries/scientific-discovery/

Schmidt M, Lipson H (2009) Distilling Free-Form Natural Laws from Experimental Data. Science 324(5923):81–85, DOI 10.1126/science.1165893, URL http://www.sciencemag.org/content/324/5923/81

Schwarz G (1978) Estimating the Dimension of a Model. The Annals of Statistics 6(2):461–464, URL http://www.jstor.org/stable/2958889

Shah M (2008) The Logics of Discovery in Popper's Evolutionary Epistemology. Journal for General Philosophy of Science / Zeitschrift fr allgemeine Wissenschaftstheorie 39(2):303–319, URL `http://www.jstor.org/stable/40390659`

Simon HA, Langley PW, Bradshaw GL (1981) Scientific Discovery as Problem Solving. Synthese 47:1–27, DOI 10.2307/20115615, URL `http://www.jstor.org/stable/20115615`, 1

Spirtes P, Glymour CN, Scheines R (2000) Causation, prediction, and search, 2nd edn. Adaptive computation and machine learning, MIT Press, Cambridge, Mass

Thagard P (1980) Against Evolutionary Epistemology. PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association 1980:187–196, URL `http://www.jstor.org/stable/192564`

Tschaepe M (2014) Guessing and Abduction. Transactions of the Charles S Peirce Society: A Quarterly Journal in American Philosophy 50(1):115–138, URL `https://muse-jhu-edu.ezproxy.lib.vt.edu/journals/transactions_of_the_charles_s_peirce_society/v050/50.1.tschaepe.html`, <p>Volume 50, Number 1, Winter 2014</p>

Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Nave Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. Applied and Environmental Microbiology 73(16):5261–5267, DOI 10.1128/AEM.00062-07, URL `http://aem.asm.org/content/73/16/5261`

Way M, Scargle J, Ali K, Srivastava A (2012) Advances in Machine Learning and Data Mining for Astronomy. Chapman and Hall/CRC, URL `http://proquest.safaribooksonline.com/9781439841747`

Woodward JF (1992) Logic of discovery or psychology of invention? Foundations of Physics 22(2):187–203, DOI 10.1007/BF01893611, URL `http://link.springer.com.ezproxy.lib.vt.edu/article/10.1007/BF01893611`

Wu X, Kumar V, Quinlan JR, Ghosh J, Yang Q, Motoda H, McLachlan GJ, Ng A, Liu B, Yu PS, Zhou ZH, Steinbach M, Hand DJ, Steinberg D (2007) Top 10 algorithms in data mining. Knowledge and Information Systems 14(1):1–37, DOI 10.1007/s10115-007-0114-2, URL `http://link.springer.com.ezproxy.lib.vt.edu/article/10.1007/s10115-007-0114-2`

Yang ZR (2004) Biological applications of support vector machines. Briefings in Bioinformatics 5(4):328–338, DOI 10.1093/bib/5.4.328, URL `http://bib.oxfordjournals.org/content/5/4/328`

Zhang NL, Nielsen TD, Jensen FV (2004) Latent variable discovery in classification models. Artificial Intelligence in Medicine 30(3):283–299, DOI 10.1016/j.artmed.2003.11.004, URL `http://www.sciencedirect.com/science/article/pii/S0933365703001350`

Zytkow JM, Simon HA (1988) Normative Systems of Discovery and Logic of Search. Synthese 74:65–90, DOI 10.2307/20116486, URL `http://www.jstor.org/stable/20116486`, 1