First published in *What if there were no significance tests?* (Eds.) Harlow, Lisa L., Mulaik, Stanley A., and Steiger, James H. 1997. New Jersey: Erlbaum. Extract. pp. 366–391.[1]

Good Science is Abductive, Not Hypothetico-Deductive

Abduction

A century ago, Charles Peirce (1839-1914) introduced the term *abduction* to denote a type of practical inference that he viewed—rightly—as importantly distinct from classically conceived deduction (logical entailment) and statistical induction.[2] This is our creation of *explanations* for observed phenomena, a mode of reasoning that not merely expands upon the information at hand ("ampliative" inference), as statistical generalization to relative frequencies in populations also does, but hypothesizes *why* these observations have their distinctive character. And these explanations invoke concepts other than the ones used to describe the observations explained. Thus from the Hartshorne and Weiss (1934) collection of Peirce's papers:

> Abduction consists in studying facts and devising a theory to explain them. (1903; Vol. 5, p. 90).
>
> Abduction is the process of forming an exploratory hypothesis. (1903; Vol. 5, p. 106).
>
> Abduction must cover all the operations by which theories and conceptions are engendered. (1903, Vol. 5, p. 414).

And most provocatively,

> The great difference between induction and hypothesis [that is, abduction] is, that the former infers the existence of phenomena, such as we have observed in cases which are similar, while hypothesis supposes something of a different kind from which we have directly observed, and

[1] (Ed.) In the earlier part of the paper, not included here, WR discussed the Hypothetico-Deductive and Bayesian approaches to scientific inference. He now turns to the third option, Explanatory Induction, which he refers to here using C. S. Peirce's term "abduction."

[2] Peirce attributed this term to a Latin translation of Aristotle's *Prior Analytica*, lib. 2, cap. 25 (Moore, 1984, p. 108). Peirce also refers to abduction as reasoning by "hypothesis" and, less often, "retroduction."

frequently something which it would be impossible for us to observe directly. (1857; Vol. 2, p. 388.)

Peirce returned repeatedly to this theme without, however, developing details on how to do it. And he urged that abduced hypotheses be appraised by follow-up tests of their additional implications insomuch as they generally remain far from a sure thing given just the observations that originally invoke them.[3]. So it might be argued that Peircian abduction was but a softer, more tolerant precursor of Popper's fiercely astringent hypothetico-deductivism, which insisted both that there is no such thing as a "logic of discovery" (i.e., that hypothesis creation is purely a psychological phenomenon governed by no principles of rationality) and that hypotheses never gain credibility save through verification of their deductive consequences. But unlike Popper, Peirce would have welcomed the prospect that some abductions have determinate forms which transmit conviction. I shall argue that such inferences are the machinery of knowledge acquisition in both technical science and everyday life.

Be that as it may, the notion of "abduction" remained quietly where Peirce left it for the first half of this century during which our discipline's method orthodoxies were codified. But the massive mid-century swing of philosophic *zeitgeist* rejecting positivistic epistemology included an awakening of philosophers' desire to acknowledge abduction somehow. Hanson's *Patterns of Discovery* (1958) was warmly received as something of a breakout despite its arguing merely that our observations are impregnated with theory from the outset while endorsing a view of abduction as the intuitive onset of not wholly implausible conjectures, without concern for any identifiable principles that might govern these. And philosophers have also become increasingly inclined to speak sagely of "inference to the best explanation" since Harman (1965) introduced this phrase as a new take on Peircian abduction, even though neither Harman nor, until quite recently, anyone else had much to say about which hypotheses are explanatory or what conjectured explanations qualify as decent much less best. However, artificial intelligence (AI) work on problem solving expressly identified as Abduction (see especially Josephson & Josephson, 1994) has begun to put algorithmic muscle into the "best explanation" slogan. Meanwhile, Simon (1973) argued in explicit opposition to Popper that there does indeed exist a logic of scientific discovery, and has since developed programmable details within his own AI framework for problem solving (Langley, Simon, Bradshaw, & Zytkow, 1987). And in Rozeboom (1961), following my personal epiphany on deriving theory from data during graduate work on

[3]Thus in the 1903 lectures: "[Abduction's] only justification is that from its suggestion deduction can draw a prediction which can be tested by induction." (Hartshorne & Weiss, 1934, Vol. 5, p. 106)

the behaviorist *What is learned?* problem,[4] I pointed out specific logical forms by which discerned data regularities are intuitively transformed into explanations thereof. Originally, I called this "ontological induction" because it creates concepts of entities (attributes and events) distinct from the ones described by the data statements. But later, in a user-friendlier and more mature statement of the position (Rozeboom, 1972), I renamed it "explanatory induction" to stake out the seminal role I claim for it in real-life knowledge acquisition.[5] Although differing considerably in detail, all these revisionist views of scientific inference fit nicely within the broad tent of Peircian abduction. Even so, the version I have called explanatory induction has a tightly operational focus warranting recognition as a distinct species within Peirce's genus.

'Abduction' is at risk of becoming a buzzword in AI circles; and the extent to which psychology's research methods can profit from study of AI data-processing algorithms claimed to be abductive is problematic. A salient case in point is the sector of AI problem solving covered by Josephson and Josephson (1994), which has a considerable history of papers with 'abduc' in their titles. This has focused on deriving from a plurality of observations on a single individual— notably, medical symptoms and lab workups in the studies published, but also potentially data clusters such as features of a crime scene, or of a handwriting sample, or of style in artwork of uncertain authorship, etc.—the best diagnosis of that individual's configuration of underlying conditions responsible for those symptoms. But the extant AI algorithms that accomplish this are preprogrammed with "domain knowledge" containing all the explanatory *if/then* laws posited to produce symptoms of the sort to be interpreted in particular cases. Algorithms that can do this effectively in difficult cases may well have applied value as expert systems; but they tell us nothing about how to conceive and acquire confidence in the explanatory laws they presume. (Elsewhere in the AI literature, programs can be found that profess to abduce laws as well; but the versions I have seen use a production logic more suited to fantasizing than to scientific inquiry.) In contrast, Simon and colleagues have said nothing in print about "abduction" or "inference to best explanation"; but their induction modeling has attempted to reconstruct discovery of laws and concepts that were historically important achievements in physical science. These

[4]What came as revelation to me was realization that although non-mentalistic S-R mediation mechanisms could in principle account for a certain prospective transfer-of-training phenomenon that commonsense would take to manifest a mentalistic "idea" mediating between external stimuli and behavior, this phenomenon would demand explanation by a certain structure of mediation with indifference to whether that structure has a mentalistic, S-R mechanistic, or some other embodiment. (See Rozeboom, 1970, pp. 120–122).

[5]My original label proves to be the superior version in that I now want to emphasize that statistical induction, too, verges upon explanation when it reaches beyond population frequencies to underlying probabilities. However, I also submit that probabilities, though undeniably theoretical entities, are not themselves explanatory mechanisms but only supervene upon those.

prestigious scientific successes are paradigm examples of the inferences I have called explanatory induction; and I welcome Langley et al's (1987) progress in simulating these as support for my still-minority thesis that interpreting observations in this fashion is the engine of advance in epistemically praiseworthy scientific belief.

Explanatory induction, or "EI" for short (not to be confused with the "AI" of artificial intelligence), has two special epistemic merits typical neither of abduction in Peirce's broadly inclusive sense nor of output from the artificial intelligence programs that have been labeled abductive by their creators: Not merely do explanatory inductions call explanations for observations to mind in the first place by decently determinate inference forms, they also yield confidence in their conclusions that in favorable circumstances can approach the vivid strength of perceptual beliefs. So why has this essay's title paid explicit homage generically to abduction rather than specifically to EI when the latter is its real focus? Three reasons: First, good science does indeed profit from broadly abductive (imaginative) speculations so long as results from innovative research provoked by these are not interpreted by blue-sky HD reasoning. Second, 'abduction' is a word that for better or worse is bound to become increasingly prominent in discussions of scientific method; so you may as well get familiar with it. And third, it is clumsy style to use an eight-syllable phrase in slogans or titles when three syllables will do.

Explanatory Induction (EI): Examples

Suppose your hand calculator—call it CP for "this calculator at present"—is acting strangely: Whenever you enter numeral 8, its rightmost display cell responds with 6, and the same occurs when the result of a calculation should show 8 in that display position. Your conclusion is immediate and assured: *Something is wrong with* CP. Were you to observe further that CP responds to entry of any digit other than 5, 6, or 8 with a non-numeric shape in its rightmost display cell never containing an upper-right vertical stroke, you would probably conclude more specifically that *the upper-right pixel at end of* CP*'s display isn't working*. But even if no digit other than 8 manifests a problem in this display position and you know nothing about CP's display mechanism, you are still confident that some feature of CP has changed for the worse even though all you yet know about this altered condition is that it makes CP show 6 in its rightmost display cell under input circumstances that formerly would have put '8' there. In either case, you have committed an act of explanatory induction on these observations from which you would have been incapable of abstaining. You have inferred that CP is in some state that degrades the desirable input/output relations it sustained previously. And you do *not* take this state to be an ephemeral manifest property that CP has just at moments when depression of a key co-occurs with a display inappropriate to that key, but which vanishes when no key is pressed or when key entry and consequent display are in agreement (e.g., when entry and right-most display digit are both 5). Rather, you feel sure that it is an *enduring* property, one that is *responsible* for some key presses eliciting incorrect responses and persists in CP even when this is resting or giving correct responses.

What you know about a properly working CP is even more richly endowed by EI. First, even if you had never received instruction in how to use CP beyond its *on/off* switch and advice that pressing CP's keys (which we will describe as "entries" corresponding to the keys' assorted labels) generally alters CP's display, you could still have become confident, through observations on CP's display changes following key entries, of many generalities having form

(1) Whenever CP is in condition C and then receives sequence K of key entries, its display at end of this input sequence is R,

wherein C specifies, among other observable preconditions, CP's display at start of entry sequence K and some information about CP's recently prior entries. In practice (and you really have acquired such beliefs about hand calculators), these generalities would for the most part accrete silently in your "background knowledge" while entering your conscious judgment as particularized anticipations of how CP's display should respond to a possible key-entry sequence initiated here and now. And the salient point to take from this is what you thereby believe about CP at various stages in a sequence of its key entries.

To become aware of beliefs that in your normal use of CP are too fleetingly transient to reach foreground attention, suppose that your execution of some calculation with CP is interrupted by a phone call. On return, you observe that CP's current display is 5 and, instead of hitting CP's *clear-all* key to start afresh, you wonder what its display would become were you to enter a few more digits, say 9, 3, 9, followed by a function key such as the one labeled '='. You realize that you can't predict CP's future display just from the present one and your choice of input starting now, but can do so if you remember enough of your transaction with CP just prior to breaking for the call. And if you are fairly sure that your last function-key entry was $\times$ while CP's display then was 1.2, you can anticipate with some confidence that were sequence 939= now to be entered, CP's display immediately following the = entry would be the product of numbers 1.2 and 5939 or, more precisely, a digital representation of that which after some paper-and-pencil scratching you can identify in advance as digit string 7126.8. But your success in predicting CP's displays in this situation isn't the point here. What matters are your beliefs (a) that CP's display R at times t' later than the present t will be lawfully determined in part by its sequence of key entries between t and t'; (b) that the present *state* of CP at t also makes a large difference for R at t'; and (c) that although this state of CP so crucial to its subsequent display is ephemeral and not directly observable, its functional character—that is, its distinctive role in production of CP's behavior—can be inferred from past input to CP and described by the conditional

(2) If the symbol string that starts with CP's current display excluding terminal dot and thereafter describes the next sequence K of CP's key entries represents a number r, and entry of K is immediately followed by one of function keys $+, -, \times, \div$, or $=$, then CP's resulting display will be standard notation for the number that equals 1.2-times-r,

to which in afterthought you append some auxiliary *if*-clauses constraining the size of r and how roughly CP is handled. Idiom urges that (2) be simplified to

(3) CP is currently disposed to respond to any input number with the product of that with 1.2,

and may even tempt you to say, more metaphorically than EI approves, that CP remembers 1.2 and is in the mood to multiply. Metaphor aside, the idiom of (3) makes it easy for you to acknowledge that when CP is working properly, it is capable of passing through many transitory states, each describable by a covert conditional—"covert" in that it is an *if/then* only implicitly—of form

(4) CP is disposed at time t to respond to input of any number r with $s \odot r$,

where s is a real number and $\odot$ is one of binary operators $+$, -, $\times$, $\div$. Since at most one completion of schema (4) is true of CP at any particular time t, (4) in effect identifies a two-dimensional array of attribute alternatives which we might call CP's "*op*(eration)-states." From there, your theory of how the s and $\odot$ facets of CP's *op*-state at any given time have been brought about by input is a straightforward inference from your generalizations of form (1).

In resurrection of the positivist program early this century for clarifying science's theoretical concepts, one might argue that CP's *op*-state at any time t is nothing more than some logical construction out of CP's key entries prior to t. But your intuitive EI leap from observations on CP to beliefs such as (4) about CP's *op*-states insists that its property described by (4) is a contemporary event which, though due to and predictable from input/output events in CP's past, is distinct from those and mediates whatever causal influence those may have on CP's next response.[6] Of course, this interpretation of (4) could be wrong. But for better or worse it is the conclusion that EI delivers here.

[6]To be sure, if you had the flu on this date last year, then it is a property of you-today that you had the flu a year ago. But we view such temporal displacements of natural events as linguistically contrived epiphenomena that supervene on the world's causal unfolding. (If you

The *op*-states of CP inferred by EI in this example are atypical of EI discovery both in their extreme ephemerality and in the accuracy with which CP's *op*-state at any given time can be predicted from its prior input history. But they usefully demonstrate the compulsive power with which explanatory inductions can occur. Given enough observations on CP to convince you that the form-(1) generalizations you have taken from these will continue to fit future observations on CP, it is psychologically impossible for you not to advance from there to indicative and counterfactual conditionals like (2) that in turn are arguably equivalent in meaning, or nearly so, to dispositional claims of form (3). These moves are far from epistemically unproblematic; but insomuch as they are as natural as breathing and nearly as indispensable, prudence advises us not to disown them but to develop expertise in their technical management.

I have taken pains to describe CP's *op*-states as "dispositions" because that is the established label for an enormous variety of qualities ascribed to things in everyday life: the sourness of lemons vs. sweetness of sugar, your uncle's stinginess vs. the generosity of your aunt, the fragility of chalk vs. the durability of wood, the stickiness of honey and adhesives vs. the slipperiness of teflon and wet ice, and so on for thousands of adjectives in ordinary language. Whatever may be the nature of what we attribute to the individuals said to be that way, it is deeply ingrained in our language that these are *dispositions*, or "tendencies" if you prefer; and over many decades of philosophers' efforts to reduce their aura of mystery (see Tuomela, 1977, for a collection of modern views) it has been more or less agreed that these are essentially the same as what we attribute by claims of form

(5) $$\text{If } S(x)\text{, then (probably) } R(x),$$

or variations and elaborations (e.g. conjunctions) thereof. In this, 'x' is placeholder for names of whatever entities we may wish to characterize this way, '$S(x)$' describes an input condition that can be imposed on x, and '$R(x)$' describes a response of x that may, though need not, be some display by another object to which x is coupled in some fashion specified in '$S(x)$'. (That is, $R(x)$ can be a meter reading.) Also, the conditionality expressed in (5) is understood to tolerate occasional failure of response R given input S. That is, whatever we mean by (5) in its disposition-demarking sense, not only does truth of '$R(x)$' not suffice for (5) to be correct, it is also allowed that (5) might hold in some instances even when '$S(x)$' is true while '$R(x)$' is false. In particular, when x belongs to a set X of entities that we think are alike in respect to (5)—notably, when X comprises a succession

have persisting health problems today, this may well be due in part to your bout of flu a year ago through the iterated dynamics of certain body changes initiated at that time; but it is surely not brought about by your having today the property of having had flu a year ago.) This may only exhibit the naivete of our causality intuitions; but it's still the way to go until the error of those becomes more apparent.

of an enduring thing's temporal stages over some limited time interval, or stages of different things that appear to be interchangeably similar in many ways—we are disposed to accept (5) if, among the members of X whose S and R conditions have been observed, R is considerably more prevalent among Xs that have been Sd than among those that have not. Indeed, in such cases we find ourselves talking about different degrees of the *If-S-then-R* disposition (which is thereby converted from an inferred state to a dimension of inferred states) such that the strength of this common to the members of an X presumably homogeneous in this respect is measured by the contrast of R's relative frequency among X-members who have been observed in condition S to its relative frequency observed among X-members lacking S.

Ordinary-language versions of such dispositional concepts usually specify their input/output conditions so vaguely and so categorically (that is, treating S and R as sloppy condition-present/condition-absent dichotomies) as to be nearly useless for scientific research save as targets of replacement by more precisely defined and more finely graded input/output alternatives. Sometimes such improvements can be viewed as more accurate diagnoses of roughly the same underlying attribute dimensions detected crudely by their ordinary-language precursors. But more often, invention of new measuring instruments—devices or special environments whose carefully controlled coupling with objects that interest us scarcely ever occurs naturally—afford display of precisely defined dimensions of quantifiably graded response alternatives that initially provide conception, and thereafter finely discriminating detection, of our studied objects' underlying properties beyond the ken of everyday experience. These previously hidden properties have now become virtually observable even if not quite so manifest as the input/output conditions from which we infer them. And data on patterns of co-occurrence among these newly discernible attributes may—or may not—lead us through iterated explanatory inductions to understanding of even deeper levels of inner mechanisms responsible for these subjects' overt behavior.

Psychology has had its full share of instrument-grounded disposition concepts, though apart from physiological psychology the "instruments" have largely been special stimulus settings with constraints on subjects' movements therein rather than the meters, gauges, chemical reagents, and other sensor devices familiar in everyday life through spillover from the physical and biological sciences. Indeed, much of educational psychology and personality assessment has been grounded on subjects' reactions to numerous brief test items—paradigmatically, choice among a small number of alternative answers to a written or sometimes spoken question—that are collected into "scales" on which the subject receives a numerical score summarizing responses to the items comprising that scale. Much has been written and still more remains to be said about the technology of such questionnaires, especially about the rationale of item grouping which in practice often illustrates

EI at work on a deeper level with techniques of factor analysis. But we cannot address that here. The point to be taken is that educational and psychological measurements of this sort are classic illustrations of explanatory-inductive concept formation accompanied by collection of data on the variables so inferred, even though their attendant circumstances seldom warrant trust at the higher confidence levels of which EI is capable, such as when these same subjects' body weight is measured by a balance or torsion scale, or their temperature by a thermometer. Each item's definition (standardized manner of stimulation and method of scoring) creates conception of a mini-tendency to get one score rather than another if so stimulated. High EI confidence that mini-dispositions so conceived genuinely exist requires subjects to show consistent differences in their responses to each item over repeated testings, which in practice is seldom demonstrated firmly. But there also exist other patterns of manifest intrasubject consistency in response to a battery of test items that urge inference to relatively enduring "traits" of tested subjects. These are measured, though with less-than-perfect accuracy, by subjects' composite scores over selected subsets of the test items and are inferred to dispose not merely subject-distinctive patterns of response to these particular test items but also—which may or may not be later confirmed—to other input conditions as well. (See McCrae & Costa, 1995.)

In practice, we seldom have much interest in dispositions identified by just one dimension of response to just one specific stimulus setting. But EI kicks in hard when we have observed a cluster of dispositions expressible in ideally simple cases by a plurality of sentences having form 'When a thing of sort B is disposed to R_i when S_id' then almost always it is also disposed to R_j when S_jd.' (The S_i envisioned here are a considerable diversity of conditions that can be imposed on B-things, each R_i, is a response made possible by input condition S_i, and B is some background condition—often conceived only vaguely by reference to a few paradigm examples—under which this consilience of mini-dispositions seems dependable.) In such cases, EI waives the mini-dispositions in favor of a single theoretical property τ whose presence/absence in a thing of kind B can be diagnosed in diverse known ways (whether the thing does R_i in response to S_i for several different S_i-tests) and is moreover expected to partake in yet-to-be-discovered lawful relations with other observable and EI-inferable variables as well. Actually, finer details of these S_i/R_i tests (notably, when some of the R_i are graded response alternatives) usually yield conception of this τ as a theoretical variable taking a considerable range of alternative values over things of kind B.

Such "cluster concepts" (as some philosophers have called them) of classic simplicity abound in chemistry, mineralogy, and medicine wherein the "natural kind" of a chemical or mineral or, in medicine, the presence/absence of a particular disease condition is diagnosed by a battery of such tests. A powerful case in point is the chemical contrast originally conceived as *acid* vs. *alkali* vs. *salt*. Partington's

(1935/1965) history of chemistry lists dozens of different tests described by Robert Boyle (a 17th Century founder of modern chemistry) wherein a sample x of some to-be-appraised material X is brought into contact with a sample s of some test material S_i. For suitably chosen S_i, changes in the appearance of s resulting from this contact reliably forecasts, for many other identified materials S_j what changes in samples of S_j will result from their contact with samples of material X. (Many though by no means all of these test outcomes are changes in s's color that depend in part on the chosen S_j. Another especially important response in some tests is s's dissolution in liquid x.) Early chemists did not learn about acids and alkalis by first speculating that such theoretical properties might exist, next deducing observable consequences of this hypothesis, and finally confirming those predictions as a triumph of hypothetico-deductive science. Rather, the alkali/salt/acid notions and their eventual refinement into a continuum of pH levels were an explanatory induction from observed patterns of reaction such as collated by Boyle. Or so I submit.

Another good example from everyday life is your learning in childhood about *hot*. When in the kitchen, or near a fireplace, or perhaps in a family workshop, you had repeated experiences wherein touching a certain object resulted in your immediately feeling acute discomfort followed a little later, if your touch was firm or prolonged, by blistering of your skin at the point of contact. You easily convinced yourself that an object which so affected you when touched would do so every time you touched it for a short duration thereafter, though usually not after some longer lapse of time. So you concluded—not by deliberated reasoning but by wired-in cognitive compulsion—that some things at certain times have an *If-I-touch-it-I'll-get-hurt* property. And additional fooling around or watching others deal with such objects also taught you that a thing having this touching-it-hurts-me property is also featured by *If it is touched by a plastic object the plastic will melt*, and by *If fat is spilled on it the fat will sizzle and likely catch fire*, and by *If a scrap of paper is held against it the paper will turn curly brown and maybe burn*, and by *If a pan of water is put on it and it stays able-to-hurt-me long enough, the water will boil.* Indeed, you learned that any one of these if/thens holding for an *it-now* pretty well guaranteed that the others were true of *it-now* as well, whence you concluded—again by innate urge though not quite so compulsively as before—that all these simultaneous *if/then*s probably manifest a single underlying condition that you came to think of as "hot" because you also learned that a nearby grown-up's shouting 'hot' when you were reaching for something also dependably indicated that your target was in this danger state. From there, you went on to discover that a thing's glowing red often signaled that it was *hot*, that devices your folks called "thermometers" can finely discriminate differences in *hot* that you could previously distinguish only coarsely, and so on for a large repertoire of beliefs about gradations of *hot* so securely rooted in your direct observations that even today you may have trouble recognizing that these beliefs constitute a *theory* about a nonphenomenal Heat variable (degrees of *hot*ness) known to you, via explanatory induction, only through its causes and effects. Indeed, commonsense is inclined to hold that heat *is* observable, just not quite so directly as some attributes we perceive.

There is one more real-science example of explanatory induction that I consider especially luminous, namely, the discovery and measurement of electrical voltage and resistance. I have cited this briefly in Rozeboom (1984, p. 220f.), but its prominence in Langley et al. (1987) encourages a more articulate review here as well. To highlight the EI essence of this achievement uninhibited by strict historical accuracy,[7] I talk about a 19th Century German physicist named Fohm, who is a mildly fictionalized version of the real George Ohm (1789-1854).

Fohm's definitive studies of electrical circuits emerged from the then-novel inventions of electric batteries and galvanometers. We know galvanometers today as magnetized needles whose rotation around a mounting pivot diagnoses the strength of a magnetic field induced by a nearby flow of electrons. But Fohm conceived of these just as detectors of a "strength of magnetic action" variable.[8] Batteries, or better, battery setups, were for Fohm special temperature-controlled assemblages of physical materials having two poles designed for attachment to wires, metal bars, or other physical objects which we shall call "loads," and were such that when a load bridged the gap between a battery's poles (circuit completion), the needle of a galvanometer held near the load showed some deflection from its resting position. Let us augment the "load" notion to include stipulation of two specific points (terminals) on the object to which a battery's poles are to be attached, so that different terminal placements on the same wire/bar/whatever define different loads, just as different temperature distributions over the same battery assemblage count as different battery setups.[9] Fohm found that when each of several different loads L_i completed the circuit separately with several different battery setups S_j, holding a particular prepared galvanometer G standardly close to L_i during its attachment to S_j rotated G's needle a consistently distinctive distance d_{ij} from its resting position.[10] And when studying his collection of d_{ij} readings, he discerned that a positive real number r_i could be assigned to each load L_i and a pair of non-negative real numbers v_j, s_j to each battery setup S_j, such that for a constant of

[7] Partly because the source material I have been able to consult is rather sketchy, but also because real Ohm's first report republished in Magie (1965) does not exhibit the EI nature of his research quite so cleanly as certain modifications of his procedure allow.

[8] See Magie (1965, p. 470)). Ohm and his contemporaries had already begun also to think of this observational procedure more theoretically as measuring a "current" of flowing electric fluid.

[9] The "thermo-electric" battery setups favored by real Ohm comprised a bar of uniform metallic composition with its poles maintained at different temperatures by iced vs. heated water. This temperature disparity was an important determinant of current production.

[10] Real Ohm's galvanometric response measure was Coulomb's more sophisticated amount of torque required to counterbalance the current-induced disturbance.

proportionality g which could be set equal to 1 by choice of G's scale unit, the errors of approximation

$$d_{ij} \approx g \times \frac{v_i}{r_i + s_j} \qquad (g = 1) \tag{6}$$

over all L_i/S_j combinations were negligible. (Henceforth, we will treat (6) as an exact equality.) These v, r, s assignments are not unique, insomuch as multiplying all the r_i and s_j by an arbitrary constant and all the v_j by another can be compensated for by a corresponding change in G's scale unit. (Tinkering with G's calibration is fair play in any case, because G's responsiveness is affected by its construction and positioning in the circuit's vicinity, and is best construed to measure current strength only up to a constant of proportionality.) But the ratios among the r_i and s_j, and separately among the v_j, are fixed by Equation 6; so by choosing a particular load L_1 battery setup S_1, and galvanometric procedure G as reference standards, Fohm was able to stipulate a unique, reproducible assignment of the v, r, s numbers in circuit experiments like this.

It is instructive to note that when Equations 6 are error-free, their right-hand terms can be determined from the data as follows: First stipulate reference-standard values for r_1 and v_1, say 1 for both. Then s_1 is computable directly from d_{11} as $s_1 = v_1/d_{11} - r_1 = 1/d_{11} - 1$ from which, for every other load L_i, we can compute $r_i = 1/d_{i1} - s_1$. Now choose an offset load L_2 whose now-identified resistance r_2 differs from r_1. Then for every other battery $S_j, d_{2j}/d_{1j} = (r_i + s_j)/(r_2 + s_j)$ so $s_i = (d_{1i}r_1 - d_{2i}r_2)/(d_{2i} - d_{1i})$ and hence $v_j = d_{1j}(s_j + r_1)$. Thus the d_{i1} for all loads L_i alternatively in circuit with reference battery S_1, together with the d_{1j} and d_{2j} over all batteries alternatively in circuit separately with reference load L_1 and its offset L_2, collectively suffice to compute all the v_j, r_i, s_j and, from there, to reproduce all the other d_{ij}. (The v, r, s assignments so identified are relative not only to choice of L_1 and S_1, but to galvanometer setup G as well. Dealing with variation in G is an additional facet of EI-grounded circuit theory that we needn't pursue here. Neither does it matter that modern data analysis would fit the v, r, s assignments by more advanced methods designed to minimize approximation errors.)

Fohm described r_i and s_j as measures of the "resistance" of circuit components L_i and S_j respectively,[11] while v_j, measured S_j's "excitatory force" (later called "voltage"). Fohm further observed that the d_{ij} readings for particular load/battery pairs were essentially constant over repeated test occasions except for some initial

[11] Initially, real Ohm took the r_i in equations (6) to be lengths of the copper wires he first used for loads in this experiment. But he also reported that when brass wire was substituted for copper, one inch of brass was current-wise equivalent to 20.5 inches of copper, showing that what plays the r-role in Ohm's law is not length of load but some other variable that is a function mainly of length among loads sufficiently alike in other respects. How soon thereafter the r-term in (6) became explicitly recognized as load resistance I do not know.

variation that might have been due to exceptionally cold weather on two deviant days. From there, Fohm and his successors were able to work out deeper laws governing voltage and resistance. Thus when load L_i, is a metal wire or other elongated object of homogeneous material with terminals at ends of its long axis and constant cross-section perpendicular to that axis, L_i's resistance equals its length divided by its cross-section area times a constant, specific to L_i's type of material though also influenced by the load's temperature, which may be called the "conductivity" of that material. Eventually, it was possible to assemble electric circuits comprising one or more current drivers and multiple loads joined by weblike interconnections and, using detectors of current flow (amperage), voltage drop, and resistance calibrated from these instruments' responses to standardizing circuits, to confirm empirically the final theoretically polished version of Ohm's law, namely, that current flow through any load in an arbitrarily complex circuit equals the voltage drop across the load's terminals divided by its resistance. Also confirmed was that a load's resistance is a property that, within limits, persists throughout its participation in a variety of circuits.

You will observe that Ohm's law and its extensions are replete with implied dispositional concepts that are not simple surface-level *if/then*s. Intensity of electric current was initially disposition to affect devices having an identified disposition to respond to magnets, but later became more richly theoretical than that. And voltages and resistances are dispositions to affect current only interactively, so that initially some of these required simultaneous diagnosis in the fashion of Fohm's experiment even though once that was accomplished it became possible to construct standardized meters able to diagnose new voltages and resistances just by a direct meter reaction. From there, a material's conductivity is uncovered as a parameter in how the resistances of loads composed of that material vary as a function of their shape, size, and terminal placement; while in all likelihood (though I have no knowledge of this) the influence of temperature on conductivity has material-specific parameters diagnostic of still deeper attributes of materials.

Three major features of EI-driven theory development are prominent in this example. First, the numbers inferred to measure dispositions affecting observed performance protrude in frugal description of provocative patterning discovered in data collected under systematically varied conditions of observation. Pairing each of n_{S} battery setups with each of n_{L} loads repetitively on m occasions yields $n_{\mathrm{S}} \times n_{\mathrm{L}} \times m$ galvanometer readings that are reproducible under (6), with scarcely any error, just from $2n_{\mathrm{S}}$ numbers $\{v_j\}$ and $\{s_j\}$ assigned to the batteries plus n_{L} numbers $\{r_i\}$ assigned to the loads. To appreciate the sense in which these v, r, s parameters characterize a *pattern* in Fohm's $n_{\mathrm{s}} \times n$ data points $\{d_{ij}\}$ (whose constancy over repetitions is an additional facet of the pattern that we shall ignore), observe that the method of computing these assignments described previously defines an algebraic function f of five arguments such that, for any fixed choice of reference battery S_1, reference load L_1, and offset load L_2, data points d_{i1} over all the n_{L} loads in circuit just with S_1, together with d_{1j} and d_{2j} over all the n_{S} batteries in circuit just with L_1 and L_2, yield prediction of all remaining $(n_{\mathrm{L}}-2)\times(n_{\mathrm{S}}-1)$ data points by formula $d_{ij} = f(d_{i1}, d_{1j}, d_{2j}, d_{21}, d_{11})$. Although this function f is easy enough to program for computer computation, it looks like gibberish if written out as a single algebraic formula. Nevertheless, it makes precise a pattern of interpredictive redundancy within data array $\{d_{ij}\}$ that when we comprehend its strength overwhelms us with conviction that these observations have common sources which plainly cannot consist in some of the G-readings being causes of others.[12] And once we appreciate the lucid elegance with which Equation 6 describes this data pattern by decomposing the surface-level f-interrelations into components associated with separable circuit parts, with the directionality in this decomposition portraying all the observations d_{ij} as similarly dependent on variables over these circuit parts whose values are estimated by the v, r, s parameters, this parsing of the data pattern demands as much a realist interpretation as do the patterns other scientists see on the output screens of telescopes, bubble chambers, and electron microscopes.

Second, the new variables defined by EI-provocative decomposition of a data pattern generally project generalizations broader than orthodox statistical generalization. The loads (and similarly the batteries) in Fohm's experiment were temporal continuants (enduring things) from which only occasional time slices (temporal stages) were in circuit with a battery. So the r_i value computed for load L_i in its temporally extended entirety actually assigned the same resistance rating to each of L_i's time slices that were in a circuit from which a galvanometer reading was taken. The empirical within-load constancy of this rating under repeated assessments urges provisional inference that within limits not yet clear, r_i is L_i's resistance (a) at every moment t at which L_i completes a galvanometer-monitored electric circuit even with battery setups not used previously, and (b) also at moments when L_i is not in a circuit at all. (Projection (a) is an empirical prediction; (b) is an instance of believing that dispositions persist even when not manifest.) And the force of projection (a) for anticipating future observations is not prediction that the past relative frequency of some observed attribute will tend to recur, but a set of conditional predictions about the behavior of new circuits containing L_i in which the conditionals' antecedents hypothesize information about the other circuit components.

Third, once data obtained under tightly controlled conditions have made such a pattern evident, it can generally be found, by systematic variation of conditions previously held constant, that this pattern's local parameters covary with other features identified (often as dispositions) in more elaborate patterns whose

[12] A thought problem for you: *Why* don't you consider it even remotely plausible that some of these G-readings, or the current strengths that dispose them, could cause the others?

additional parameters disclose still more source variables. Thus had Fohm's experiment suffered large temperature fluctuations from one circuit reading to another, or (contrary to physical possibility as we know it) his loads asynchronously changed their metallic composition every few hours, it would have been nearly impossible for him to discern any lawfulness in his data. But once pattern (6) was perceived under tight constraints and confirmed by replications, it become straightforward to study how resistance varied as a function of conditions previously held constant and discover conductivity in the parameters thereof associated with a load's material.

Explanatory Induction: Overview Principles

The generic character of explanatory induction can be summarized as follows:

1. For a scientific data collection to have any interpretive significance, it must exhibit some regularity that cries, or at least softly pleads, for generalization to other events.

2. When a generalizable regularity observed in local data is described with quantitative precision (generally an idealized pattern qualified by goodness-of-fit ratings), its description includes certain parameters that we suspect depend in part on local conditions that will change when the form of this regularity recurs for data of this sort in other situations.

3. When parameters of a local-data pattern are found to vary with changes in local background constraints that have been selectively relaxed with some care, that dependency will likely manifest a pattern having parameters of its own, and so on for an iteration of parameters disclosed by constraint relaxations.

4. When working observed pattern parameters into stories about what seems to be responsible for what, stories that we need to guide our predictions of new events from our accumulated experience with these phenomena, we are often compelled—willingly—to treat those parameters as estimated values of hidden variables[13] that we know only through the data regularities they dispose. However, data patterns are often not so clear as we would like;

[13] Source variables inferred from covariations have become commonly known in the factor-analytic/structural-modeling literature as "latent variables" or "latent factors." This label is somewhat misleading, insomuch as "latent" ordinarily connotes inactivity, whereas were hidden variables inactive we could not detect them. But once we do detect them, they are not really hidden anymore either. What we need here, but have not yet found, is a deft adjective connoting "observed only indirectly." (All observations are indirect to some degree, but some rather more so than others.)

even when strong they may be frugally parameterizable in more ways than one; and systematically controlling or passively observing additional factors ignored previously may change the pattern's gestalt.

To illustrate this last point, which is important, consider the research of Rohm, an alter ego of Fohm whose experiments more closely resemble those of real Ohm. Initially, Rohm used a single battery setup S_1 and, repeatedly over several days, took the G-readings of S_1's circuit completions with eight copper wires $\{L_i\}$ having the same rectangular cross-section but large differences in length. Each load's G-ratings were averaged over repetitions of L_i's circuit with S_1 to define mean datum d_i for this load. And when each d_i was inverted to $y_i = 1/d_i$ and plotted against wire length x_i over all these loads, it was plain that y was a linear function of x with a positive slope b and additive constant a. Rohm next observed G-readings on circuits of S_1 four brass strips with the same cross-section but differing in length, and found the dependence of y on x in this set of brass loads to be again linear with the same additive constant but a different slope. Next, he changed the battery setup to S_2 and found the same linear data pattern as before within each same-metal group of loads, but with the pattern's a and b parameters both altered by the battery shift. However, the proportionate change in b under shift in load-metal with the same battery was constant across batteries; so Rohm was able (the algebra is simple though not entirely obvious) to summarize all these results by a law just like Equation 6 except that Fohm's r was replaced in Rohm's version by $m \times x$ for a numerical constant m specific to the load's metal. Rohm could have viewed m at this point as measuring a conductivity property of metals. But he aborted that conception when his study of varying the cross-sectional areas w of his loads showed that $m \times x$ equaled cx/w for a deeper numerical parameter c that differed across load metals. Rohm's c is the same conductivity coefficient that Fohm discovered when, after identifying load resistance, he looked for determinants of r. But Rohm's pattern parameters did not explicitly exhibit r at this stage of his research. Only later, when he varied loads over irregular shapes and odd terminal positionings did Rohm perceive that it was most insightful and computationally universal to characterize loads by their r-terms in pattern parameterization (6). That a load's resistance could also be estimated from its physical dimensions and the conductivity of its metal in the restrictive case of cylindrical loads with terminals at their ends had importance for practicalities of electrical transmission but mattered little for basic circuit theory.

Explanatory inductions are thus provisional conclusions that may well become revised, expanded, or superseded altogether by more intricate configurations of factors, hidden or manifest, as more of the total pattern enters our ken. But that does not mean that explanatory inductions are usually *wrong*, anymore than it is wrong to say that John weighs 150 lbs. when his exact weight is 153.2 lbs. Human cognitions seldom if ever match reality exactly. But not all cognitive inaccuracies are equally fallacious. Getting things roughly right is often all we need for the purpose at hand; and assertions that are somewhat vague or not entirely correct can still be vaguely or approximately true.[14] Just like material technologies, our conceptual resources and the beliefs these enable are constantly evolving, due in no small part though by no means entirely to science's explanatory inductions. But modern replacement of slide rules by electronic calculators doesn't show that slide rules didn't do good work in their day; and neither does the likelihood, that much in our contemporary repertoire of EI-driven theoretical constructs will eventually grow obsolete, at all impugn their value in affording the truest account of natural reality currently available to us.

Epilog on Practicalities

As you have observed, this essay has not attempted to advise you on specific statistical techniques to favor when personally analyzing and interpreting data. Rather, it has surveyed three major contrastive outlooks[15] on the logic of scientific reasoning with intent to promote certain attitudes that should make a difference for what you do in research practice. In broadest outline, I encourage:

1. Feel free to draw upon imaginative speculations when planning a research study or soliciting support for its execution, but resist with all your might, so far as supervisors, grant appraisers, and journal editors allow, actually interpreting your resultant data as a binary pro/con vote on some hypothesis proclaimed in advance. As a corollary, shun null-hypothesis tests while appreciating that the statistical models these exploit readily avail you of conclusions vastly more informative than pass/fail grading of H_0.

2. Sniff at Bayesian confirmation models with cautious welcome, like a guard dog appraising strangers in the company of his master. You will seldom be positioned to revise any of your unconditional beliefs in accord with its quantitative ideal, nor would trust in the outcome be wise if you could. But when working toward a deliberated conclusion of some importance to you, the quality of your reasoning may well profit from a dry-run attempt to simulate Bayesian derivation of this conclusion from beliefs in which you feel considerable conditional/unconditional confidence, so long as you translate Bayesian conditional-Crs into probabilistic *if/then*s that you can verbalize.

[14] Failure to develop accounts of truth and reference that allow semantic aboutness relations to be matters of degree rather than all-or-none remains a large blemish on modern philosophy of language.

[15] (Ed.) The three outlooks are Hypothetico-Deductive, Bayesian, and Abductive. See Note 1 above.

And although I have not developed the point, there are special contexts of data interpretation, notably ones involving inference to and from probabilistic generalities, where the Bayesian model offers guidelines through the fog in which commonsense intuition abandons us.

3. When analyzing data, try to summarize these as conforming to some predictive regularity describable by a small number of parameters within an orderly form such as exemplified by algebraic equations of modest complexity. (By "predictive" I mean that under this regularity, appreciable parts of the data array can be reproduced with decent accuracy from its remainder.) When circumstances permit, it is desirable to perceive this local data pattern as a fragment of patterning more strongly evident, and perhaps more intricate, in the collation of your results with findings by other studies varied in their conditions of observation. And when interpreting that pattern, view these parameters at least provisionally as measuring features of your data's underlying sources. Moreover, when seeking to replicate this pattern—an exceedingly important phase of research practice—under variation in the background constraints whose local constancy has enabled this pattern to become detectable, expect these parameters to emerge as variables that are predictively interrelated with other variables whose local constancy has also been relaxed.

Unpersuaded? Then let me offer another triplet of recommendations, not particularly rooted in what has gone before but focused on how to choose statistics for analysis of sample data.

THE STATISTICAL RELEVANCE PRECEPT: *Make sampling statistics your servant, not your master.* When designing an experiment, or analyzing data already in hand, first of all ask yourself what summary features of these results you would most desire to learn, or would like to see replicated if you have noticed them already, were this data array's sample size so large that any statistics you choose to take from it have essentially zero sampling error. Next, compute the values of these statistics for your sample data and think on how you would provisionally interpret them, either in cognitive inferences or in decisions to act, were you confident that they were population values. Finally, work out some conception of how this putative knowledge of your target population is degraded by the sampling error that statistical theory warns must in fact contaminate your sample statistics to some degree, and think through how that should attenuate your interpretation of these results.

Comments. This precept is neutral on conflicting philosophies of inference. It is your right to highlight a statistic because its value is predicted by a hypothesis this experiment is designed to test, or even, if you insist, to collapse its continuum of metrical alternatives into two or three categories. But if your primary interest shifts adventitiously to a feature of your results not anticipated by any statistical model you had intended to apply (e.g., when a bivariate relation's trend that your preplanned model describes by a linear correlation coefficient appears strongly curved in the data), that too is your right. The precept is also neutral on what may be meant by your statistics' "population values," in particular whether these (a) are abstractions from your data variables' joint frequency distribution in a larger totality of real individuals sampled by your subjects, or (b) are features of a probability distribution over these variables, conditional on a complex P of background properties common to your subjects, that is indifferent to how often P is realized. (These construals of "population" don't really need airing just yet; but your elementary statistics training has so often encouraged you to presume the first view that a caution against complacency in your understanding of this notion seems warranted.)

The precept's second admonition is more devious. It doesn't urge you to draft the journal article or lab report you would produce were your sample size truly enormous, but invites you to try on a frame of mind that frees your statistical thinking from enchantment by the sampling-theoretic tail that far too often wags the data-interpretive dog. Your struggles to master the technicalities of textbook statistics may well have depleted your zeal to question which outputs of their procedures are what you want to learn from your data, especially if you rely on data-analytic computer programs that squeeze numbers out of your raw input by routines whose documentation of algorithm or motivation you do not fully comprehend. Someone has judged these outputs relevant to questions they considered worth asking. But which of them matter for *your* questions? (How long should you treasure the t statistic computed for a mean difference in your data, or search for meaning in the numerical values of ANOVA's F ratios for a sample trend's main and interaction components in preference, say, to graphic comparisons of group means?) You have nothing to gain from concern for a statistic's sampling uncertainty (save to oblige colleagues who want it) if you have little idea of what to do with its population value were you to know that. Beyond that, if the statistics you have selected do indeed seem right for your intended interpretation, how precise a determination of their population values would your interpretation be able to exploit? (Might information just that they lie in one broad region of possible values rather than another be all the detail that you can use?) The point of asking is not to justify a needlessly sloppy report, nor even to take comfort in current tolerances for imperfect accuracy, but to think on this experiment's position in the larger scheme of work underway on its topic. If you don't know how interpretation of your chosen statistics could profit from learning their precise population values, shouldn't you be trying to develop a conception of your inquiry's target that does?

The precept's third admonition prompts two closing considerations, one tactical and the other strategic. Tactically, left to your own devices you would undoubtedly be at a loss to put much precision into your spread of subjective uncertainty about the population values of statistics you have sample-estimated. But through the logic of deriving particularized *if/then*s and consequent *since/therefore*s from probabilistic generalities, sampling-theoretic *confidence intervals* for your sample-statistics' population values give you near-unconditional posterior credibilities which, after some fine-tuning for which you may feel need,[16] may well be about as good as your rationally disciplined uncertainty about population statistics can get. To be sure, such posterior statistical confidences are conditional on the idealized premises of the statistical model under which they have been computed. And my fine-tuning qualification acknowledges the broader point, about inference to particulars from probabilistic generalities, that the strength of your *c-is-B* belief channeled from your *c-is-A* conviction through the *if/then* conduit set up by your confidence that $\Pr(B \mid A) \approx r$ may well be modulated by other things you also believe about object c. Alternatively, if fine-tuned confidence intervals are not to your taste, you can estimate Bayesian posterior credibilities for your statistics' population values by computational procedures detailed in the technical literature. Except when confirming results from previous samples from what you are confident is the same population, this should seldom differ enough from your spread of uncertainty framed by confidence intervals to make any practical difference.

Strategically, however, efforts to be precise in our uncertainty about a sample statistic's population value are often pointless. The reason is a dirty little secret: We seldom have much clue to the identity of this population or for that matter any good reason to think of it as unique. According to statistical theory, probabilities in a mathematically structured system thereof are always conditional on one or another configuration P of population-defining properties variously described as preconditions, background constraints, local constancies, or other phrases similarly connotating limitation. In principle any simple or complex property can play this role: Whenever we envision a probability distribution $\{\Pr(A_i)\}$ over an array A $= \{A_i\}$ of attributes, we can always think of some constraint P such that each $\Pr(A_i)$ is really $\Pr(A_i \mid P)$, whereas for any A in $\mathbf{A}$ the restriction of $\Pr(\ )$ to $\Pr_A = \Pr(\ \mid A)$ defines what can be treated as an ostensively unconditional probability distribution over the more restricted population $P{\cdot}A$. When the probability calculus is applied to sample data comprising an array of attributes distributed over a set $\mathbf{s}$ of subjects, its "population" provision is interpreted as some conjunc-

[16]Notably, when the interiors of some confidence intervals include regions of the real-number continuum wherein the statistic at issue cannot lie, for example, negative values for variance ratios.

tion P of properties, common to all the members of $\mathbf{s}$,[17] which statistics jargon treats as a population from which $\mathbf{s}$ is a "sample." But enormously many different populations are sampled by $\mathbf{s}$: In addition to common properties $\{P_i\}$ expressible by various logical compounds of the predicates used to describe subjects in our data records, the members of $\mathbf{s}$ share numerous additional properties $\{P'_j\}$ that we could have put on record but didn't, and uncountably many more $\{P''_k\}$ that we couldn't have ascertained at all, especially the ones of which we currently have no conception; and every conjunction $P_i \cdot P'_j \cdot P''_k$ of these is also a population sampled by $\mathbf{s}$. So when we talk about population values of statistics computed from the $\mathbf{s}$-data, which population do we have in mind? This wouldn't matter if our sample statistics at issue had the same population value in each of these, but of course that is wildly untrue. Is there something sufficiently distinctive about one of these to qualify it as *the* population relative to which $\mathbf{s}$-sample statistics have population values? The most logical candidate, the conjunction of all $\mathbf{s}$-shared properties, is a nonstarter; for that includes being a member of $\mathbf{s}$, leaving no room to generalize. Another contender, motivated by sampling theory's yen for "random" sampling, is the P from which $\mathbf{s}$'s sampling is most nearly random. But even if we had a concept of random sampling that is not largely vacuous,[18] it would not define an *effective* criterion by which we can plausibly judge, from all we know about some other individual i not in $\mathbf{s}$, whether i too is in this population most randomly sampled by $\mathbf{s}$.[19]

[17] "Common to all" doesn't preclude fine differences in what is said to be in common: For example, the members of a group of preschool children all share the property of being younger than six years (or some other upper bound) even when their exact ages are all different.

[18] The problem with this notion is not so much *random* as its coupling with *sample*. A random *variable* is just a scientific variable (dimension of attribute alternatives) over whose values we have posited a probability distribution conditional on some background condition. And in its original conception, a "random sample" is a set of things picked from a larger set $\mathbf{S}$ of real objects by a procedure under which each subset of $\mathbf{S}$ of given size has the same probability of selection. But what could it mean to sample randomly from a population-defining *property* that is indifferent to the prevalence or identities of its instances? The best prospect for extending the original selection-from-$\mathbf{S}$ sense of this to an open population P is to say that a set $\mathbf{s}$ of P-things is a "random" sample from P if the procedure by which $\mathbf{s}$ has been picked imposes an additional constraint C such that the probability distribution in $P \cdot C$ of certain distinguished variables has special stipulated features (e.g., equiprobability of all their values). But this prospect is highly programmatic: It would at best yield many different types of random sampling whose motivation remains obscure; and in any case, to my knowledge neither this nor any other approach to defining random samples from open populations has been advocated in the statistical literature. Old-style random sampling from extant totalities continues to have applications value—for example, in demographic surveys, quality-control batch testing, and assigning research subjects to treatment groups—but its scope is quite limited.

[19] For simplicity, I have put the population problem in terms of properties common to all individuals in a data collection's subject sample $\mathbf{s}$. But preconditions of generalization also come in more complicated versions that incur correspondingly more complicated indeterminacies of population identity, especially when distributions of independent variables and statistics conditional on particular values thereof are at issue. To a large extent these can with some artifice be treated

Let's approach this problem from another direction. Given our sample statistics and an inventory of all the properties we know to be shared by our sample subjects **s**, how do we judge whether a statistic's value computed from the **s**-data should well-approximate the value of that same statistic in another group $\mathbf{s}'$ of individuals whose scores on the relevant variables are still unknown? In practice, we base such judgments on the similarity of properties shared in $\mathbf{s}'$ to the ones shared in **s**. But we don't include all the latter in this appraisal because in the first place we don't know most of them, and secondly don't think that all the ones we do know are relevant. We believe—never mind why, though there is much to say—that some **s**-shared properties made a difference for the **s**-data's statistics whereas others didn't matter, and that if $\mathbf{s}'$ too shares all the former then, regardless of what happens in $\mathbf{s}'$ on the latter, the statistics observed in **s** should well-approximate the corresponding statistics in $\mathbf{s}'$ so long as sample sizes are respectable. But we are also pretty sure that relevant/irrelevant here does not coincide with known/unknown. So even if most of the known **s**-shared properties are also common to $\mathbf{s}'$, statistics from **s** may not generalize to $\mathbf{s}'$ very well if $\mathbf{s}'$ lacks some of the **s**-shared properties relevant to those statistics, whereas these statistics may generalize nicely to an $\mathbf{s}'$ lacking some **s**-shared properties if those happen to be irrelevant. How we manage to learn which features of our data's background constancies make a difference, and what differences they make, are matters for a treatise on experimental design, not closing thoughts. The closing point to be taken here is simply that it makes little sense to fret over how closely our sample statistics approach their population values until we have made an honest effort to say *what* population, and to identify that in terms of properties we have reason to believe really matter. Meanwhile, you can't go wrong by heeding *Steiger's Maxim*:

> An ounce of replication is worth a ton of inferential statistics. (Steiger, 1990, p. 176).

---

as special cases of the simple version. For example, each value of an independent variable is a property common to a subset of **s** that samples a subpopulation of any population sampled by **s**. And by viewing **s** as a single entity of which our individual sample subjects are parts, we can include holistic properties of **s**, such as the frequency distribution of an independent variable, in the defining features of a population of sets from which **s** is a sample of size 1. Even so, these more elaborate sample features contributing to population indeterminacy warrant explicit recognition because they underlie important operational issues of research design, notably an experiment's "controls."

References

Hanson, N. R. (1958). *Patterns of discovery.* Cambridge: Cambridge University Press.

Harman, G. (1965). The inference to the best explanation. *Philosophical Review*, *74*, 88–95.

Hartshorne, C., & Weiss, P. (Eds.). (1934). *Collected papers of Charles Sanders Peirce* (Vols. 1–5). Cambridge, Mass.: Harvard University Press.

Josephson, J. R., & Josephson, S. G. (1994). *Abductive inference.* Cambridge, England: Cambridge University Press.

Langley, P., Simon, H. A., Bradshaw, G. L., & Zytkow, J. M. (1987). *Scientific discovery.* Cambridge, Mass.: MIT Press.

Magie, W. E. (1965). *A source book in physics.* Cambridge, Mass.: Harvard University Press.

McCrae, R. R., & Costa, P. T. J. (1995). Trait explanations in personality theory. *European Journal of Personality*, *9*, 231–252.

Partington, J. R. (1935/1965). *A history of chemistry* (Vol. 2). London: Macmillan. (Originally published 1935)

Rozeboom, W. W. (1961). Ontological induction and the logical typology of scientific variables. *Philosophy of Science*, *28*, 337-377.

Rozeboom, W. W. (1970). The art of metascience, or, What should a psychological theory be? In J. R. Royce (Ed.), *Toward unification in psychology.* Toronto: Toronto University Press.

Rozeboom, W. W. (1972). Comments on professor Wilson's paper. In J. R. Royce & W. W. Rozeboom (Eds.), *The psychology of knowing.* New York: Gordon & Breach. (pp. 390–398)

Rozeboom, W. W. (1984). Dispositions do explain; or, picking up the pieces after Hurricane Walter. *Annals of Theoretical Psychology*, *1*, 205–223.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*, 173–180.

Tuomela, R. (Ed.). (1977). *Dispositions.* Dordrecht, Netherlands: D. Reidel.