First published in *The Journal of Clinical Psychology*, *64*, 2008, 1109–1127.

The Problematic Importance of Hypotheses

Abstract

Contrary to what seems to have become established dogma in the behavioral sciences, the confirmational relevance of some data possibility D to a conjecture H should be indifferent to which of these two prospects we think of first. In particular, if ascertaining whether D is true would be a Popperian test of hypothesis H were D's implication by H declared at outset, the confirmational relevance of D to H is not diminished by learning D before becoming aware that H implies it.

Prelude

While some still-problematic aspects of scientific reasoning obdurately resist deeper clarification, certain others do not. I shall introduce this attempt to advance our grasp of an oddly neglected facet of the latter by splitting it from a huge overhang of the former. And be assured that the generic importance of hypotheses is not in question. The issue is how we should adjudicate them.

Surely we can agree with only modest bicker over phrasing that "science" is *professionalized pursuit of empirical knowledge*, wherein "professionalized" means advanced expertise, not skill for hire. The pivotal notion in this definition is "knowledge," which for present purposes needs some unpacking despite being, like most of the big concepts we live by, too elastic for tidy definition. Indeed, there are at least three importantly distinct varieties of "knowledge" (ambiguous in English but clearly distinguished in German as *kennen* vs. *können* vs. *wissen*—see Rozeboom, 1972b), all of which are enhanced by extant sciences. But foremost of these is *propositional knowledge*, which is something that we can communicate by declarative sentences. (Ideally, "propositions" are the meanings of those. How often and to what extent the declarative sentences we utter in real life approach semantic ideality, or how philosophy of language should explicate that ideal, is far out of range here.) There exists considerable philosophic consensus that propositional knowledge is *justified true belief*, but details on what each word in this compound means are still disputatious and it is neither practical nor necessary to recapitulate my surveys thereof in Rozeboom (1967, 1972a). What's salient here is that knowledgeable belief comprises some of the propositions believed by you, me, and others in our intellectual communities (elsewhere too); and that as scientists we are professionally committed to producing and promulgating importful propositions in our areas of expertise that we justifiedly believe to be true at least approximately. We don't often succeed at this, of course: Most of a live science's cognitive activity comprises efforts to verify or refute (i.e. justify an extremity of belief/disbelief in) currently interesting propositions whose truth-status is elusive. But as you well know, even partial success in that can profoundly transform both our intellectual and material environments. The point of this introduction, which may seem pretentiously vacuous, is to make plain on record that the primary professional mission of a modern science of some topic T is to engineer beliefs about T-matters. The same can be said about political and commercial-products propaganda except for the crucial difference that whereas the latter is designed to influence by any legal means where its recipients position these assertions on the optative-mode axes of their propositional-attitude space,[1] scientific communication seeks to induce beliefs only at whatever levels of conviction are sustainable by plausible argument from consensual evidence. But articulate metalogical accounts of what nondeductive inference forms warrant what degrees of credence transmission from their premises to their conclusions still remain nearly nil; with the result that although scientific instincts (read critical commonsense) continue to yield progress that has become torrential in the physical and biological sciences, analysis and interpretation of data in the behavioral sciences has been stunted by canonizing the Popperian hypothetico-deduction model of scientific advance in our still-sacrosanct ritual of statistical null-hypothesis testing. NHST's blurred vision has been decried often enough of late to need no further protest here; it is the more generic relentless demand for hypothesis testing that needs reappraisal. Still prefatory to that, however, is some expansion of the modifier in my "nearly nil" disclaimer above.

As you surely know, there does exist one articulated theory of rational belief change that is operational in some idealized circumstances and admonishes how we should try to reason from evidence to inconclusive conclusions even when our uncertainty about our prior uncertainties precludes serious respect for any posterior credibility ratings we might compute. This is Bayesian confirmation theory, whose onset was a 1764 posthumous publication by English Presbyterian minister Thomas Bayes that seems to have initiated (somewhat cryptically—see Dale, 1991) the notion of inverse probability. When developed into to a view of rational belief strength positing a measure Cr (for "credence") over propositions that satisfies the postulates for statistical probability while ranging from 0 for certainly false to 1.0 for certainly true, this implies how strongly the Cr-value of a hypothesis h should be increased by obtaining evidence e that is more plausible if h is true than if it is not. The algebra of Bayesian confirmation can be found in many sources nowadays, starting with Google returns for "Bayes' Theorem" and, close to home, several overviews of Bayesian inference in Harlow, Mulaik, and Steiger (1997). Its

[1] For a brief introduction to the multidimensionality of propositional-attitude space, see Rozeboom (1972b, p. 38)

qualitative core is the metabelief that when a data-possibility d is implied (not necessarily with deductive certainty) or at least not contradicted by a hypothesis h, the change in h's credibility most rational when verifying or refuting d enhances our store $\boldsymbol{k}$ of prior knowledge is a determinate function of (a) h's prior credibility $Cr(h \mid \boldsymbol{k})$ for us just before we learn whether d, (b) d's prior credibility $Cr(d \mid \boldsymbol{k})$ (for us, etc.), and (c) what $Cr(d \mid h \cdot \boldsymbol{k})$ would be for us if h could be added to our knowledge store.[2] $Cr(d \mid h \cdot \boldsymbol{k})$ equals 1.0 (maximal) if d is a logical consequence of h-and-$\boldsymbol{k}$, and can also be computed for lower Cr levels when h includes adequate statistical assumptions. But such extreme confidence doesn't seem metarational when we aren't aware that h entails d given $\boldsymbol{k}$; and more generally, our inability in research practice to proffer plausibly reasoned values of $\mathrm{Cr}(d \mid \boldsymbol{k})$ and $\mathrm{Cr}(h \mid \boldsymbol{k})$, is a nearly insurmountable obstacle to deriving a numeric posterior probability $\mathrm{Cr}(h \mid d \cdot \boldsymbol{k})$ for h in light of our new data.[3] Even so, attempting to think through how we might identify or at least decently approximate these elusive prior credibilities might well expand our comprehension of possible explanations for the phenomena at issue more deeply than we get just from learning whether d.

Regardless of how feasible Bayesian reasoning might become under exceptional circumstances, mere consideration of that possibility calls attention to an awkward untidiness in human cognition, scientific or otherwise. This is simply that our beliefs, no matter how justified, are mental[4] states that wax and wane in our thinking from moment to moment in response to fluctuating sensory input and other antecedent neural events. But many of our momentarily active beliefs presumably derive—causally—from enduring attributes that dispose belief arousal by present cues of diverse sorts both external and internal. The immediate practical payoff of this, of course, is enabling perceptual input from our current situation to activate beliefs that guide our choice of weal-promoting actions. But the conceptual content of a recalled belief seldom if ever precisely reproduces the prior propositional ideation whose memory trace has disposed this recall (or at least

[2]The dot here is standard logicians' notation for conjunction "and," while the vertical bar can be verbalized as "presuming" or "given." For simplicity I have spoken of background propositions k here as "knowledge" when in real-world application it would comprise antecedent beliefs generally failing to qualify as that by tough epistemic standards. This illustrates why, when discussing science's epistemic pursuits, we should try not to speak too glibly of what we "know."

[3]The pivotal Bayesian posit is that our synchronic credences for any two propositions p and q should satisfy $Cr(p \cdot q \mid \boldsymbol{k}) = Cr(q \mid p \cdot \boldsymbol{k}) \times Cr(p \mid \boldsymbol{k})$, which yields that any one of these three credences is a determinate function of the other two. It does not, however, specify which should influence what in our reasoning processes. In particular, $Cr(q \mid p{\cdot}\boldsymbol{k})$ isnt required to result through an inferential conduit from p to q; it can just as well derive from our $Cr(p \cdot q \mid \boldsymbol{k})$-to-$Cr(p \mid \boldsymbol{k})$ ratio regardless of how we get those credences from our background knowledge $\boldsymbol{k}$.

[4]"Mental" by our commonsense conception thereof in ordinary language without presuming an immaterial ontology for them. "Identity" theorists including myself consider mentation to be aspects of the same brain events we conceive by other routes of conceptual access as physical. Mental vs. physical is arguably a distinction more of semantics than of ontology and some version of pan-psychism remains a serious possibility.

we have no good reason to think otherwise); and even more problematic—since to my knowledge (which could well be deficient on this point), modern cognitive psychology has never explored this—is how closely the credence given to a proposition when contemplated previously is included in its storage trace and reproduced in this belief's recall. And that in turn raises a provocative question: Suppose that proposition p is true and my previous active believing-that-p was justified then in whatever manner and strength is needed to qualify my cognition then as a knowing-that-p. If some time has passed during which p-believing had faded from my active awareness but some current cue has now reactivated my contemplation that-p, do I still know that-p? Arguably, often not for two reasons: First, the cued reactivation of my p-thought may not have restored my previous confidence that-p, and it seems passing strange to claim that I still know that-p even though I am not at present sure of it. And secondly, if my prior belief-that-p was a knowing-that-p by virtue of adequate justification, is it still a knowing if later cueing reactivates my belief that-p but not my justification for so believing? For example, if I know at this moment that there are exactly four single-digit prime numbers because I still remember the trivial proof of this I have just thought through, do I still know that tomorrow when I recall this proposition but not the proof? What if I don't recall my justification but do remember that I had one? And speaking of memories, when our main support for believing that-p is our remembering that-p, under what special circumstances does that adequately qualify as knowing-that-p when memories are demonstrably so unreliable? The salient point here is not that a precise definition of "knowledge" remains elusive but that our belief strengths and metabelief justifications of them need repeated recreation. Our recalls of these those compare less to retrieving steel instruments from our mental toolbox than to horticultural germination of seeds from prior harvests.

1st Movement

What has just been overviewed is the "huge overhang" my opening remarks have scheduled for evasion. The personalized instability of justified believing makes acquisition and retention of knowledge seem nearly impossible - which is to say, considering this tortuous cascade of uncertainties, that it might be prudent to downgrade claims of "scientific knowledge" to "scientific plausibilities." But there is still the sociological facet of science's epistemic enterprise to consider. By that I mean not just interpersonal interactions whereby one person's verbal output directly influences the propositional thinking of others, but also - far more importantly - production of physical communication objects which enduringly retain stimulus configurations that can evoke, in persons suitably exposed to them, propositional thoughts largely the same as those they cued in the persons who issued them. I am, of course, envisioning hard-copy linguistic communication broadly conceived as any form of physical mediation whose enduring structure is imposed by articulated thought in one person and elicits counterparts thereof in suitably receptive others that often include later stages of its creator. Visual communication by written words is foundational; but graphs, charts, and numeric displays are also visual communicators, and within the past century storage and dissemination of auditory ideation transmitters has also become pervasively effective. The importance of such archived collective memories (for simplicity, call these "documents") can scarcely be overestimated, but you are already well aware of that.[5] I introduce this theme here partly to preface its appearance in later conclusions but also to point out that although durably accumulating document troves vastly enhance the range of propositions we can contemplate beyond those to which our personal brain traces give us recall access, and activate imported thoughts with visual or auditory word-string cues far more vividly articulate with on-demand persistence than afforded by our intra-brain recall, it generally delivers these propositions with even fainter credence than internal recall supplies. (By "faint" here I mean not active disbelief but little actionable confidence at any level, such as when we read fiction.) So for archives to convey knowledge or at least some attainable proximity to that (if we can't reach justified certainty, highly plausible in light of strong evidence will do), its primary-proposition documentation needs to be accompanied by secondary documentation that cues beliefs about the truth-relevant circumstances of the primary documents' creation and storage. Most salient therein is description of evidence alleged to support the primaries; but claims about how these documents were created and appraised by others in our comunications environment, modulated by our assessment of their epistemic credentials, are also cogent. There is vastly more to be said about this documentational facet of science, with development of that taking a shape rather like what we see when looking into a mirror while facing that with another held just below our eyes. But that is overhang to disregard at present. Aspired here is merely some mitigation of the damage done to our harvests of research information by preaching to students that "scientific method" is Popperian hypothesis testing made scientifically precise by focus on statistical hypotheses, and enforcing that primacy by publication strictures. My overview thesis—which by rights should be so commonsensical that calling it mine is like claiming credit for contention that icy sidewalks are slippery—is

> **Motif**: If our becoming confident that p (for simplicity, call this "learning" p without presuming that such learning usually qualifies as know-

[5]Popper (1972) has developed this theme at considerable length with commendable emphasis on its importance. But in his wont of carrying good ideas to counterproductive extremes he classified this reservoir of belief prompts as "objective knowledge." Not a whole lot of knowing goes on in library archives, because most of the propositions expressed by words written there aren't true with certainty if at all and the books presumably don't do much thinking of them either.

> ledge acquisition) confirms to some degree the credence we give to the possibility that q, it shouldn't matter whether our learning p occurs before or after our initial becoming aware that q is possible.[6]

Obviously this thesis needs some explication, especially on what complications are to be partialled out by "shouldn't matter." First, let's stipulate that whether q is true is value-neutral for us and not influenced by our actions.[7] (Those are important considerations for a treatise on applied epistemology, but unhelpfully distractive here.) And we want to exclude egregiously irrational belief changes which, though of course not an affliction of our own thinking, does prompt us to appreciate that our intuitive reasoning vastly transcends the inferences approved by established deductive logic. In order for learning p to confirm q for us to some degree (whose grades include indifference and disconfirmation—negative confirmation—as well as enhanced support[8]), several things must occur in our conscious(?) thinking, where "?" acknowledges that much of this may elude our introspective awareness and in retrospect be considered "intuition". Plainly, ideas p and q need co-activation (though not perfect synchrony), each coupled with a current degree of belief (feeling of assurance). And we must also experience some connection between propositions p and q in virtue of which the credence we feel in p influences our q-opinion. Such belief modulation is far less perspicuous than its prevalence in our conduct of daily living might tempt us to presume. Even in ordinary instances of simple deductive logic, e.g. when we are sure that we left our car keys either in our coat pocket or on the kitchen table, and from failure to find them in our pocket become hopefully confident that they must be on the table, intellectual recognition that p entails q isn't at all the same as becoming sure of the latter as a result of learning the former. Even so, developing a modicum of linguistic sophistication enables us to observe logical relations among propositions

[6]It shouldn't need saying, but better be said anyway, that 1st-person-plurals here and elsewhere in this essay dont refer to everyone, but foremostly to myself and readers whose thought processes cohere with mine. In this venue that should be common; but elsewhere less so.

[7]The non-influence qualification here needs clarification, since in experimental research we generally start with an inference of form If h, then contriving conditions c will have results r. What is to be free of experimenter influence is not occurrence of c but how prospect r turns out given c. But this raises a point of considerable interest though ferally digressive here: If we take If c then r to be h's unconditional implication to be tested, it becomes arguable that h can be confirmed simply by withholding production of c. To defeat this conclusion, we need to deny that its If __ then ... is the extensional implication of modern logic which translates that as Either not-__ or ..., and posit that what we test in cases like this is the expanded hypothesis h *&* c. But then we have the problem—trickier than it may first seem—of deciding how a shift in our *Cr(h & c)* should alter our *Cr(h)* when we are seldom positioned to feel sure that c has been achieved essentially as wanted. (See the Appendix for more on the obscurities that arise for inferences from the outcomes of if/then predictions.)

[8]You should, however, continue to understand occurrences of "confirm" in this essay to mean positive confirmation unless context clearly calls for the broader construal.

of concern to us established by words they have in common and feel consensual belief influences among these in ways induced by connectives and quantifiers in the sentences expressing them. But in many if not most instances where we consider ourselves to be reasoning deductively, it is not so much an experience of p's conceptual connection with q that confirms q from p in us as it is some metabelief that becoming sure of p should so influence our q confidence. For example, raw memory or written notes may remind us of having previously observed an inferential relevance of p to q even though we don't feel the confirmational force of that just now. And more importantly we often believe that p confirms q because we are aware that a document we respect asserts this, especially when that includes detailed arguments for the claimed implication. Such inference-by-authoritative-hearsay is a mainstay of modern quantitative data analysis wherein positive or negative confirmation of explanatory hypotheses by sample data depends strongly on statistical theorems and numeric tables whose derivations few of their users could fully understand even if they were game to try, and also requires us to trust that our computer printouts correctly implement our wanted application of that mathematics.

Envisioning a credenced p, q, and credence connection *CC(p,q)* between them becoming roughly co-active[9] in our thinking positions us to consider whether the order in which these thoughts are activated should matter for the belief changes that ensue when our brains' rationality dynamics work to harmonize their credence levels. Despite prodigious efforts by philosophers of knowledge, logical deduction is the only specific form of *CC* yet understood well enough to guide our reasoning praxis with justified confidence. Normatively, this advises us that whenever we feel sure that p logically entails q (and hence also that *not-q* entails *not-p*), we should never believe p more strongly than we believe q nor doubt q more than we doubt p. Finding ourselves guilty of that irrationality urges us to revise these belief-strengths, but generally leaves us unsure of what adjustments are most compatible with our current credence distribution over the rest of our belief system. The handbook on management of that has yet to be written.

However, just a paragraph or two in that much-to-be-desired but little-to-be-expected handbook should suffice for our present preamble point. First, consider its deductive extreme: If proposition p logically entails proposition q and we become actively aware, perhaps at separate times, that p and q are interestingly possible, how should the credence we give to one of these influence our credence in the other? By rights, our *Cr(q)* should be no less than our *Cr(p)*; but failure of

[9]A detailed account of real-life confirmation would likely admit if not insist on some microasynchronies among these. Beyond a generic posit of asynchrony between causes and their effects, which wants evidence to precede confirmation, I have no specifics to proffer on this. Fortunately, they don't seem needed here.

that is not obviously irrational if we don't discern that $p \to q$.[10] Possibly in this case, subconscious brain processes might ensure that $Cr(p) \leq Cr(q)$ results in us just from our active thinking both p and q without need for additional awareness of $p \to q$; but I know of no evidence for that, either introspective or psychonomic. Our prior p-thinking and q-thinking may well have had cumulated effects on our levels of *Cr(p)* and *Cr(q)* at onset of our occasions of $p \to q$ appreciation, but surely what matters when inference occurs is only our Cr-values active when the inference channel is opened, not the history of how they arrived there. Be that as it may, it would be clearly irrational to feel sure that p entails q and yet believe q less strongly than p.

This special case of strict entailment also illuminates how consequence verification induces confirmation, at least if we accept idealized confirmation theory's premise that for a rational thinker entertaining propositions of high semantic quality[11]—preconditions henceforth presumed until later relaxed—degrees of credence have a metric, Cr, that can be scaled for convenience to range from 0 for utter disbelief to 1.0 for complete certainty. When considering the logical bearing of proposition q on proposition p, we can split the credence composition of that required for coherence as

$$
\begin{aligned}
&\text{(i)} & Cr(p) &= Cr(p \mathbin{\&} q) + Cr(p \mathbin{\&} \textit{not-q}) \\
&\text{(i*)} & &= Cr(p \mathbin{\&} q) \quad \text{if } p \text{ logically entails } q \\
&\text{(ii)} & Cr(\textit{not-p}) &= Cr(\textit{not-p} \mathbin{\&} q) + Cr(\textit{not-p} \mathbin{\&} \textit{not-q}) \\
&\text{(iii)} & Cr(p) + Cr(\textit{not-p}) &= 1
\end{aligned}
$$

with similar decompositions of *Cr(q)* and *Cr(not-q)* that don't matter here.[12] Suppose that we have provisionally set credibilities for the conjunctive propositions on the right in (i–iii) that satisfy these epistemic norms. How to reach this cognitive state for a q that did not occur to us when our interest in prospect p first arose is

[10] The arrow denotes logical entailment, as common though not universal in logicians' notation.

[11] Classical deductive logic presumes that the sentences it governs are semantically ideal in ways our real-life locutions generally fail to perfect. Three major suboptimalities still not adequately comprehended by philosophy of language and logic are vagueness, ambiguity and, for nominals, nonunique reference (i.e. multiple or none). All three are illustrated by "Pegasus was a clumsy flier," wherein "clumsy" is plainly vague, "flier" is ambiguous among "stays airborne by wing flapping," "travels by airplane," and "comprises one or more pages of distributable announcements," and "Pegasus" originally had no real referent but now, I suspect, may refer for some people to one or another toy or pet. This last prospect also points to another linguistic imperfection, namely, that the same word or phrase is often understood by different people or the same person at different times to designate different things.

[12] A partition like this of a proposition p's momentary credence conjoined with one or more other propositions epistemically relevant thereto might be called a "credence-space assay" of p. We don't need this locution here, but it helps to manage a rough patch in the Appendix. It also abuts other epistemic problems with disjunctions that don't protrude here, notably the Pandora's Box of how these differ from abstractions.

a serious issue which, however, would be unhelpfully distractive here. (Unrecognized entailments are conception suboptimalities differing in kind from vagueness and ambiguity but similarly troublesome for the philosophy of knowledge.) But how to adjust credences in light of deductive consequences has been and remains a foreground issue of applied epistemology, and (i*, ii, iii) makes plain why consequence verification should be confirmatory: When the Cr-levels on the right therein are set with $Cr(q) < 1$,[13] and some input increases *Cr(q)* to 1.0 (though weaker confirmation of q also works similarly), *Cr(not-p & not-q)* in (ii) drops to zero while its former value needs to be additively distributed between *Cr(p & q)* and *Cr(not-p & q)* in order to keep (iii) satisfied. Bayesian confirmation theory implies that this redistribution should preserve the ratio of those; but even if we have qualms about the simplicity of that ideal, it remains extraordinarily difficult to imagine epistemic circumstances under which all of this transferred credence could rationally be added just to *Cr(not-p & q)* with none for *Cr(p & q)*. And if the latter gets any at all, *p & q* and, by (1*), hence p is thereby confirmed at least a little.

The sketch just given of *Hypothesis first, Evidence later*, doesn't plainly support my thesis that which comes first shouldn't matter. Indeed, if "confirmation" narrowly construed means credence increase induced by evidence, initial evocation of a conjecture by evidence that supports it doesn't plainly qualify as confirmation. But it does cohere with contention that whether awareness of a hypothesis comes before or after awareness of what is eventually appreciated as evidence for that shouldn't matter for the ensuing degree of support. Yet "shouldn't" isn't co-extensive with "doesn't," so let's consider a situation wherein we experience a series $\langle e_1, \ldots, e_i, \ldots, e_n \rangle$ of high-credence beliefs (potential evidence in want of explanation) that eventually induces us to think of a hypothesis h that, had we conjectured this prior to our string of e-belief acquisitions, would have been strongly confirmed by that. Is our *Cr(h)* when we first become aware of possibility h only after learning these e-facts the same as what it would have become then had we conjectured h at outset of our e-findings? Not necessarily, for several complications: First, just continuing to bear possibility h in mind throughout the period of e-acquisitions without noticing their h-relevance might well give h some credibility unless some obviously disconfirmatory evidence intrudes. (For casually considered possibilities, familiarity breeds content.) Second, if the evidential import of e_i for h isn't applied when this evidence is fresh, its force may well be degraded by suspicions of unreliability when retrieved from memory or external archive for later appraisal of its h-relevance. Third, actively reappraising h throughout the sequence of e-considerations may sufficiently enrich our comprehension of the ideas in h that conceptually h is no longer the same theory at end of its e-inquisition that it was at outset even if its verbalization hasn't changed. Fourth, conscious

[13] If $Cr(q) = 1$ at outset, q can't be confirmed further.

concern for prospect h from the outset of e-accumulation may well make us passively receptive to if not actively productive of additional evidence bearing on h which we obtained only because h's possibility occurred to us before all the e-data were in hand. Finally I venture, entirely by data-bereft speculation, that experienced trial lawyers could advise us how some sequences of evidence-and-argument presentation are more effective than others for persuading a jury to return the verdict wanted.

Haven't I just discredited the "overview thesis" proclaimed so grandly at outset here? Not at all. That was brandished canonically to initiate contention that confirming conjecture h by verifying its previously inferred consequence e rationally supports h no more than does learning e first and discovering afterward that h implies it. But this initial declaration admitted that there were complications to address; and beyond the specifics on that just mentioned, any appreciable interval between our *Cr(h)* at one time and our acquiring evidence e for or against h at another is an open port for other cognition-influencing inputs that also may modulate what our *Cr(e)* does to our *Cr(h)* when those get together in our active thinking. So the Thesis may well be an oversimplification; but the dogma on hypothesis appraisal it is proffered to de-enfranchise is even more so.

2nd Movement

Even so, isn't hypothesis appraisal's problematic sensitivity to the timing of evidence delivery a philosophy-of-knowledge issue far too esoteric to concern empirical science practitioners? Apparently not, at least not in the behavioral sciences. For quite some time introductory-psychology texts have standardly included a section on "scientific method" which is generally stated (with some variance in wordings and emphases) as hypothesis-testing steps:

1) Make some observations that invite explanation.
2) Devise a hypothesis that implies what we should find if we make more observations of this sort under certain stipulated conditions.
3) Collect some data that way that way and determine whether or not this prediction is successful.
4) If it is, you could be onto something and may publish your results; if not, revise your hypothesis and try again.[14]

[14]This view's spread is considerably broader than psychology texts, though I have surveyed that expanse only through Google returns for "scientific method." One nicely concise phrasing of this (by Jose Wudka, MIT Professor of Physics) to be found there is: "The scientific method is

There is nothing particularly reprehensible in this, but it makes formalized hypothesis testing the engine of scientific progress with no suggestion that some findings of the Step 1 sort can be just as epistemically productive as any Step 3 result and often more so. And amplifying that bias, behavioral science's enthrallment by the mystique of modern statistical theory has long been training our students to believe that report of any empirical research result worth communication must include appraisal of its sample statistics in light of some hypothesis about the data's explanatory sources embodying a posited probability distribution for their observable output. Stir in the notion that a hypothesis test isn't authentic hypothetico-deductive theory appraisal unless the hypothesis is explicitly conjectured prior to the test results, and we reach the opinion that when a substantive-hypothesis-*cum*-sampling-disturbances posit h fits a highquality dataset well within plausible sampling distance of what h implies should be found absent sampling error, h's credibility is appreciably enhanced by this fit if h was stipulated at outset to be the hypothesis tested, but earns little if any credit if h has only been recognized *post hoc* as the best fit within an identified set of alternatives to h.

Scorn for *post hoc* model fitting has been especially strong in the structural modeling sector of modern behavioral science.[15] In an internet discussion forum

---

the best way yet discovered for winnowing the truth from lies and delusion. The simple version looks something like this:

1. Observe some aspect of the universe.
2. Invent a tentative description, called a hypothesis, that is consistent with what you have observed.
3. Use the hypothesis to make predictions.
4. Test those predictions by experiments or further observations and modify the hypothesis in the light of your results.
5. Repeat steps 3 and 4 until there are no discrepancies between theory and experiment and/or observation.

When consistency is obtained the hypothesis becomes a theory and provides a coherent set of propositions which explain a class of phenomena. A theory is then a framework within which observations are explained and predictions are made."

[15]Note for readers unfamiliar with modern structural modeling: Originating in the path-analysis of geneticist Sewall Wright, this has become a hypothesis-intensive expansion of inferential factor analysis. Given the covariances in some data sample among suitably many metric variables, a structural model thereof hypothesizes that these derive from their subjects values of unobserved (latent) source variables that discernibly effect these data variables through some diagnosable pattern of import for those. Structural modeling goes beyond classic common-factor analysis in conjecturing a network of causal paths through which the latents can influence not only the data variables but one another as well. These causal connections are modeled as algebraic equations that are linear in their causal-strength parameters; a particular model is selected by stipulating (inter alia) which direct path connections have zero strength (i.e. don't exist); and model solution consists of finding values for the open parameters that maximize a measure of the accuracy with which the model-reproduced data covariances approximate approach as subsets of all different ways to stipulate values for some of these parameters, and in principle a suitably powerful computer can be programed to solve enormously many of these alternative path structures for best-fit coefficient values and corresponding accuracy with which these reproduce the empirical

for this that I monitor, authorities regularly admonish neophytes that significance-test appraisals of post-hoc models chosen in light of fits better than that of the source structure conjectured at outset lack the probative import of statistically acceptable fit by a pre-hoc model because such revisions "capitalize on chance." I submit to the contrary that an excellent fit discovered post hoc by scanning numerous alternatives to whatever structure may have been hypothesized initially deserves as much respect as would have been appropriate had it been chosen at outset to be the hypothesis officially tested. The only qualification needed is that the stronger is any pre-hoc belief we may have conceded to a model discerned post-hoc to yield excellent data fit, the stronger is the resultant credence this deserves as well.

To solidify support for this thesis, here are some quasi-realistic hypothesis-testing scenarios that should damp any dissonance remaining between that and your own intuitions about intelligent data interpretation. Suppose that as head research director in a foundation for socio-economic studies, you have contracted with a consortium of marketing agencies to explore what psychological determinants—tastes, values, skills, beliefs, aversions, aspirations etc, and principles of their acquisition/comport/satiation—most predictively control the market choices of consumers. You and your staff associates, Andy and Barb, know of many conjectures that have been aired on this matter, some in its extant literature, others your own, and together more than enough to enable your team to devise many observational measures of these putatively relevant human traits and control some of the external conditions that influence them. Substantive details on these don't matter here; it suffices for you to envision yourself in a setting that affords an abundance of numerically scaled data, some presumably derived causally from your subjects' locations in a multi-dimensional space of psychonomic-attribute alternatives (values of inner-state factors) while other data dimensions scale external input conditions that you suspect have helped to position these subjects in this inner-state space. You are seeking to develop one or more hypotheses yielding equations of structural-equations modeling (SEM) form

$$\mathbf{Y} \approx H_{\mathrm{y}}(\mathbf{S}, \mathbf{X}), \quad \mathbf{S} \approx H_{\mathrm{s}}(\mathbf{S}, \mathbf{X})$$

wherein H_{y} is an algebraic function that maps each sample subject j's array $\mathrm{X_j}$ of observed-input scores (if any) together with j's inner-state coordinates S_{j} into approximately j's array $\mathrm{Y_j}$ of observed output scores that more or less imperfectly manifest j's inner-state properties, while H_{s} comprises functions hypothesizing how some of the observed input and/or inner-state variables affect others in causal processes that work their way to $\mathbf{Y}$. [If you are acquainted with structural modeling , and maybe even if you aren't, you don't need the rest of this paragraph.] When interpreting Y-data by SEM, our posited hypothesis is a specification of multivari-

data relations.

ate functions $\langle H_\mathrm{y}, H_\mathrm{s}\rangle$ (usually linear) containing open parameters whose possible numeric values are construed to scale strengths of causal influence along paths hypothesized by the SEM equations' structure, with pre-assigned zeros demarking path connections allowed by the generic model form but posited null in this application. The hypothesis tested is mainly that the Y data under analysis have unobserved sources and/or mediators ("latent" variables) that connect the Y and maybe X variables in accord with this model's path structure; but it also includes idealized presumptions about the shape of the source variables' joint distribution in a population from which the subjects actually observed are supposedly a random sample.[16] And the observational consequence of this structure-*cum*-distribution hypothesis that can become confirmatory evidence for it is the existence of numeric values (which model solution can then find) for the initially unspecified pathweights under which the data covariances implied by this model fit match their actual values in this subjects sample closely enough for some currently orthodox statistical rite to consider this discrepancy "statistically insignificant."

Here are some scenarios for how you and your research team might react when fits of such models bring into play your opinion of post hoc confirmation:

> **Scenario 1**: Prior to collecting sample data D, you explicitly proclaim model M_1, which embodies currently prevailing views on the determinants of market choices, to be the hypothesis tested. Subsequent best fit of M_1 to D yields large errors of data-covariance reproduction, so you advise your sponsors that their current marketing strategies may be degraded by false assumptions which should be correctable, if they approve continuation, by results of another study. When they do, your team formulates a different model M_2 which eliminates the more egregious inaccuracies in M_1's best fit to D, and prepares to collect more sample data on which to test its fit by M_2.

Comment: One tiny qualm aside, this is surely the machinery of modern scientific research running smoothly and productively as it should.

But the qualm, which I encourage you to share, is not really tiny. No one should dispute the desirability of seeking confirmation of M_2 before this project's sponsors base large changes in their marketing strategies on it. But they certainly deserve an alert that you have obtained evidence which challenges presumptions of their current practice that they may well want to modify if your present discovery proves to replicate robustly. And they can't start operational contingency planning for that unless you advise them not only that you have discredited M_1 but also that M_2 seems much closer to the way things actually are. Arguing that M_2 hasn't

[16] See Rozeboom (1997, p. 386ff) on the near-universal deficiency of our efforts to identify the populations we presume our data to have sampled.

gained any credibility from your study because your team didn't think of M_2 as a possibility at outset, much less infer from it then any prospective D-features that if found would confirm it, would call into question your epistemic competence.

> **Scenario 2**: Prior to official stipulation of M_1 as the model to be tested, your appreciation that this may not be completely correct urges that your team prepare to learn more from your forthcoming D than just whether M_1 should be rejected. Accordingly, while you yourself officially predict that D will satisfy M because all three of you agree that past evidence favors this, you also instruct your two associates each to think of a different structure, perhaps with a different number of latents but not flagrantly implausible, that they can respectively sponsor as their specific hypothesis to be tested by the forthcoming D. So Andy and Barb come up with different alternatives M_a and M_b to M_1 that are formally elegant though rather too much so to seem very plausible; and after D collection followed by solving each model for open-parameter values yielding best fit to D, you are all surprised to find that the best-fit $\mathrm{M_b}$ solution, unlike the M_1 and M_a fits, reproduces the D covariances well within conventional tolerance for sampling perturbation.

Comment: Different collaborators sponsoring different models should seem appropriate if you believe, or think is prevailing doctrine in your larger research community, that a hypothesis h can be confirmed only by verifying some data possibility d inferred from h and proclaimed as a test thereof prior to ascertaining the truth of this d. But it raises questions about what requirements, if any, extant hypothesis-testing doctrine puts on pre-test belief and post-test credence broadcast. Is it permissible for your colleagues to pick the models they respectively sponsor just because these were simple or perhaps conceptually intriguing despite seeming implausible? (I strongly suspect that this often occurs in real-world educational settings where students learning to do multivariate research or meet requirements for a higher degree therein must produce hypotheses to test.) And if you publish your verification of a prediction deduced from some hypothesis you worked up only because you needed an easy one on which to practice scientific method, should your surprising result be taken seriously by others? (Unless you are suspected to have faked your data or bungled their analysis, why not?) More immediately, since Barb predicted $\mathrm{M_b}$ but Andy and you backed alternatives that failed, is it acceptable for you two losers now to favor $\mathrm{M_b}$ with as much credence as Barb's entitlement? (If not, why?)

> **Scenario 3**: The same as Scenario 2 except that because you and your colleagues differ in your interpretations of extant data and past theoriz-

ing on this topic, you also disagree on what model for the forthcoming D should fit best. You expect to verify M_1, but Andy and Barb are willing to make small wagers on M_a and M_b, respectively.

Comment. In this case, your team members' respective choices of hypothesis to test are not arbitrary: Laying wealthexchange bets on the test outcome authenticates that you three have genuine personal credences associated with these hypotheses, and how this experiment's results have changed those can be diagnosed from how they affect your betting on a replication test. If your initial betting was $10 each in a pot won by backing the model that reproduces the D covariances most accurately, and you agree to bet with the same choices as before but with larger stakes on which of these proves best in your replication study, wouldn't you and Andy insist that Barb should put more into the winner-take-all pot than do either of you? This scenario could be expanded with betting on your replication's outcome in a larger group of your professional peers who have received progress reports on this project. (Never mind that in real life the proprietary nature of this research would disapprove such early broadcast.) If they create a betting pool on your forthcoming replication study, do you think they should agree to equal odds on which of your team's three models best fits the new data; and would you expect their consensual betting odds (hard to establish but no betting absent that) be appreciably affected by knowing how strongly M_1, M_a, and M_b were believed by the various members of your team prior to learning their respective D-fit accuracies? (I could suggest special conditions under which that would be rational, but they would verge upon science fiction.) The point of Scenarios 2 and 3 is that learning how well a given model fits a given dataset does make a difference for the credence we give to it, and should do so pretty much regardless, if the data collection and analysis is competent, of who believed or predicted what at start of that, albeit credence prior to that input also matters for the posterior credence that results.

Scenario 4: Same as Scenario 2 except that your associates and you jointly think of several possible models including M_1, and make the disjunction of this set your joint prediction.

Comment. To my knowledge, testing a set of hypothesis alternatives simultaneously has seldom if ever been explicitly approved in our methodology literature. Yet that is precisely what we do when solving a model with open parameters for best fit. Intuitively, there is an ontological difference between channels of causal connection among the loci of events and the qualities of influence that pass through these,[17] so that positing the former should be choice of a model while solving for its

[17] See Rozeboom (forthcoming) for conceptual tools to recognize causal displacements more explicitly in our model formalisms.

parameters informs us about the latter. But the partially ordered nests of mathematical functions that define alternative model forms suggest no such contrast, and no other criterion has been established for distinguishing model structure from outset specification of a pathweight subset. So it's hard to discern much rationality in partitioning the totality of a multivariate function-form's parameters into a subset left open for optimizing model fit while the rest are frozen at prespecified point values. Accordingly I submit as my generic thesis for SEM, that when analyzing data D collected to test a hypothesis H, if we discover that a different model H* whose structure had received little if any consideration at this study's outset has open-parameter settings that reproduce the data moments more accurately than does the best fit of H, the first-time credence this finding imparts to H* should in principle exceed that which remains with H. I say "in principle" because if H has been well-supported by previous studies which D largely replicates, you have every right to suspect that the superiority of H* in this instance may be a sampling fluke. In practice, if you take H* seriously but further tests of that will take considerable preparation and grant renewal, you would want to check how well it fits other datasets that previously supported H if some can still be accessed. And you can run a bootstraps study of D to see how consistently one of H and H* outperforms the other in that. But my take-home point here is simply that giving cautiously appreciable credence to a model whose good fit to D is discovered post hoc is no different in kind from the cautiously appreciable credence you give, say, to confidence intervals computed from D by your favorite sampling model for the open parameters in the H you have chosen pre hoc for testing.

In short, I am trying to convince you, should you be among those who have been conditioned to think otherwise, that what is cogent in the hypothetico-deductive model of scientific progress is simply its being a useful way to motivate and focus pursuit of evidence relevant to uncertain generalities whose truth matters for us. Our lives teem with belief influences; but the only form of nondeductive inference that has firm meta-logical justification is confirmation of conjectures by verifying their decidable consequences,[18] whence it has been cogent to point out that sciences gain epistemic stature by promoting this as their prime inquiry procedure. But when hypothesis h is confirmed (positively or negatively) by our arriving at an extremity of belief in some decidable e entailed by h, what rationally matters for this inference impact is only our becoming aware (an experience that can then be stored in memory and external archive for later revival) that e is a consequence of h and is true (or false). There is no inherent relevance in which part of this joint awareness was activated first; that only affords the complication that if h has been under consideration prior to arrival of e's import, it will bring to this

[18]Note that apart from the Bayesian inference model, it still isn't clear even how justification of confirmation by consequence appraisal can be stretched to approve our use of modern sampling statistics to achieve confirmation by verifying consequences that don't follow with certainty.

epistemic encounter a degree of *Cr(h)* different from and probably more hardened than what it would be were our first thought of possibility h to be evoked by our awareness of e. Neither needs the credence you give model M in light of its fit to data previously archived in a research publication be influenced by additional information about how that data induced its producers to adjust any opinion they may have had of M unless they are authorities from whom you take your own beliefs about topics to which M is relevant. In that case, pending replication of your results you might consider deferring your own provisional judgment of M to what you think is theirs.[19]

Coda

Clarifying how affirmation of predictions should influence the credence merited by hypotheses which imply them is assuredly worth metascientific concern but should not divert our attention from far more serious obscurities in the scientific pursuit of knowledge. The floodgates restraining those are burst by appreciating that when hypothesis h entails evidence e, the alternatives to h that also do so are a diversely infinite subset of all the assertions constructable in our language; and while demonstration thereof is just a prefatory near-triviality, what remains when that is brushed aside is rough terrain still inadequately explored. Seemingly minor is that if, for any proposition k logically consistent with the h in logical entailment $h \rightarrow e$, expanded hypothesis h *&* k also nontrivially[20] entails e and is confirmed by verifying that consequence. But many eligible k are disconfirmed by e, yet could have been implicitly if not explicitly part of h from the outset were this adventuresomely conceived—which demonstrates once again (cf. Rozeboom, 1972a, p. 101, Rozeboom, 1997, p. 337ff., and more generically the Appendix below) that when h entails e, even though verification of e confirms h as a whole it may well fail to confirm and can even disconfirm some of the constituent posits conjoined in h. Yet our primary practical incentive for confirming a scientific hypothesis is to build trust that other observation-language inferences we draw from it (notably, of form If-we-do-this-then-that-should-result) are also veridical. Just as falsifying some of a theory's predictions doesn't establish that everything in that is wrong (a point that post-Popperian philosophy of science has taken pains to acknowledge), neither does verifying some of its predictions merit increased confidence in everything the theory implies. Introducing fledgling students to "scientific method" by the 4-stage

[19] By rights, we should expand at this point on details of the multifaceted complexity, scantily surveyed near outset above, of factors that influence the *Cr(h)* we manage to activate for adjustment in light of new evidence. But Rozeboom (1997, pp. 346ff) has already attempted that at greater overview length than feasible here, and I encourage you to pick up on that for debate. Much remains to unfold.

[20] Were k inconsistent with h, this conjunction would be logically false and degenerately entail all propositions.

testing cliché is probably harmless if not morphed into the null-hypothesizing statistical ritual; but it's important for those who advance to graduate science to learn that holistic accept/reject statistical-test appraisals of a target theory is no more creative science than warming a frozen supermarket dinner is gourmet cooking. An alpha-grade theoretical scientist should be able to search out and devise diagnostics for basic features in a probated theory's conceptual structure that haven't yet been linked to observable consequences, and to rejoice when new procedures of data production disclose hitherto unknown data patternings that deepen our abductive access to the explanatory sources of these phenomena. When we bring evidence to bear on a hypothesis warranting the close scrutiny motivated, say, by fear that getting its import wrong could bring on a technological disaster (uncontrolled chain reactions, lethal virus mutations, ring tone supersonics that harden eardrums, etc.), and the steps of inference from hypothesis to observable consequents can be made verbally explicit, there is an operational procedure that can clarify just what components of this hypothesis matter, and how, for its various data implications. Execution of this diagnostic requires a modicum of concept-analytic skill that isn't taught in science methodology courses; and since I have already outlined the process in Rozeboom (2005, p. 1344–1349), I needn't rehash it here. It suffices to admonish that astute evidence appraisal focuses on select features of the hypothesis at issue with only secondary confidence adjustments, if any, in its remainder. Holistic acceptance/rejection is for amateurs. (For more on this theme, see the Appendix.)

Finally, a word about interpreting data patterns. In neither the behavioral sciences nor modern philosophy of science has it been adequately appreciated how our most firmly established theoretical concepts are sustained not by deliberated hypothetico-deductive reasoning but through natural abductive inferences from the characters of data patterns to concepts of entities (mainly attributes) taken to be their explanatory sources. I describe these as "natural" because they are of a kind with our statistical generalizations, whose problematic justification has exercised philosophers of knowledge ever since Hume's sceptic challenge. But short of infinite regress, not everything we build our lives upon can be explained or justified; and natural selection has gifted us with both statistical and ontological (explanatory) induction as hard-wired survival processes that we take to be rational, though not infallible, with no more need for metarational justification than we have need to justify our breathing. Both induction forms continue to be largely successful, we have no reason to expect that to change, and you can read expositions of explanatory induction in Rozeboom (1972a, 1997) richer than a few summary sentences could proffer here. As thoughtful SEM practitioners occasionally point out, although statistical testing of a model's fit to data can confirm a joint probability distribution for the datascores in a population from which these have been sampled, and suggest commonalities of their determination by underlying sources, this yields little if any increased support for conjectures about the substantive natures of those. Yet there do exist discovery-friendly research procedures within the purview of standard SEM applications that tentatively yield such information, and explanatory induction is the inference framework which develops them. You have been practicing explanatory induction daily throughout your entire life (well, probably not in utero); and attempting to suppress it when seeking scientific explanations of empirical phenomena would be like mandating that we should never eat fresh fruit lest we might sometimes bite a worm. Post-hoc data interpretation is a glory of empirical science, not a sin.

Encore

Don't shun interpretation of unpredicted data patterns; revel in the opportunity.

Appendix A:
A Tested Hypothesis's Haze of Variations

Although it is generically a sound principle of hypothesis appraisal that verifying some previously uncertain consequence c inferred from a hypothesis h increases h's credibility, the precise logic thereof is considerably murkier than seems adequately appreciated in our methodology literature. Even if pitfalls of data interpretation potentiated by this obscurity seldom(?) seriously degrade our research conclusions in practice, some explicit recognition of their potential should facilitate our management thereof.

One odd cluster of problems for the hypothetico-deduction model of scientific inference arises from *hypothesis inflation* and *consequence dilution*: If hypothesis h entails consequence c (abbreviate this as $h \rightarrow c$) while a and d are additional propositions such that $a\&h$ is not logically false nor $c\text{-}or\text{-}d$ logically true (that is, ignoring trivial extremes), then also $a\&h \rightarrow c$ [h-inflation] and $h \rightarrow c\text{-}or\text{-}d$ [c-dilution]. These seem to imply that confirmation allows us to make dubious conjectures plausible by appending them to some theory generating a strong track record of successful predictions, and to diminish prediction failures of theories we much want to succeed by disjunctive cushioning of their more dubious implications. And though commonsense surely shields us from inferences so egregiously perverse, we still want some metatheory clarifying how such arguments go wrong and some concern for whether subtle versions of these may not sometimes degrade our real-life reasoning.

Hypothesis inflation is as common as food mold and often not merely benign but operationally unavoidable. Almost always in scientific research and scarcely less frequently in everyday life, our hypothesis-based deductive predictions take the form "If h, then c always occurs in circumstances a," or more briefly, "If h, then c-wherever-a." (Inference this ideally determinate is in practice usually softened to probabilistic import. But problems of strict entailment rationality don't vanish from probabilistic inference; they just get murkier there.) When empirical research on prospect h finds or experimentally contrives an instantiation i of observable condition a, whether i also manifests c is a test of inflated hypothesis $a(i)\&h$ which, however, focuses the confirmational impact of its success or failure just on h since $a(i)$ retains its truth presumption (though possible uncertainty about that complicates our story here). But an h with nontrivial real-life implications can almost always be analyzed as a conjunction of constituent propositions; and h's confirmation as a whole by HD consequence verification seldom if ever confirms all conjunctive constituents of h equally. Indeed, it may even disconfirm some, especially in cases where competing hypotheses all predict c.

An important admonition for science praxis follows immediately from this: **Never interpret results of a hypothesis test holistically**. For your research to advance our understanding of the topic addressed by hypothesis h, it is nearly worthless for us to learn simply that your test of h has confirmed/disconfirmed this by verifying/refuting h's data implication d. Unless h is trivially simple, it is logically a conjunction of many propositions, not all of which are needed for h to entail your declared prediction. Those which are not inflate a more austere portion h^* of h sufficient to imply d, and have no manifest claim to any of the belief change warranted by our learning whether d. So far as you are able—and, like athletic dexterities, this skill appears not to come easily to many—you should try to identify components of h that can be expunged from h without impairing the deflated h^*'s import for your study's results. It does not, however, follow that whenever h is logically equivalent to $g\&h^*$ with h^* alone sufficient to predict your experiment's d-outcome, our g-credence should be indifferent to that result. If g has explanatory import for h^*, confirmation of h^* by your d-finding should also pass some confirmation back to g. And if we suspect that d may be over-determined, i.e. has multiple sources perhaps including g that can bring about d even absent h^*, our current belief repertoire may approve an adjustment of our g-credence in light of your d-finding independent of how that affects our h^*-belief. ("Our" in the preceding sentence goes proxy for members of our scientific community. Broader sharing of belief updates is more problematic.) It is seldom practical for a research report to say much about all the explicitly or implicitly identifiable propositions whose credibilities are to one extent or another differentially affected by a hypothesis test, but it's important to appreciate that its h is generally a conflation of many ideas and needs separate appraisals of its most salient parts.

Consequence dilution, on the other hand, seems to be more a metatheoretic curiosity than an operational threat. Considering that in general, the more abundantly hypothesis h entails diverse consequences that prove true the more sure of h we become, it appears that once we learn that h implies d, we can generate a diversity of additional propositions $e_1, \ldots, e_i, \ldots$ now uncertain but soon to be determined (e.g. predictions of weather and imminent competitions in sports and politics) to yield disjunctive consequences $\{d\text{-}or\text{-}e_i,\ i = 1, 2, \ldots\}$ of h, each of which by rights should confirm h if its e proves true. Once we learn whether d, these dilutions no longer matter since if d is found to be true its confirmatory blowback is arguably the same as that of d's totality of logical consequences, whereas learning that d is false also falsifies h beyond any redemption by other consequences of h that are true. But in many circumstances we can learn the truth of arbitrarily many e_i in this dilution set while d remains uncertain; and whereas false e_i shouldn't much threaten h so long as d remains possible, ones that are true not only confirm h but do so cumulatively by verifying the corresponding $d\text{-}or\text{-}e_i$. Or so it seems.

It is intuitively plain that something must be wrong with this argument; the problem is to diagnose just what. It is not that disjunctive predictions are illicit. These often occur in scientific practice, e.g. when a data parameter is predicted to lie within a specified numeric interval. More importantly, for any hypothesis h of form "*All As are Bs*" and individual i that might have attribute A, h apparently entails "*Either B(i) or not-A(i)*" which can quickly accumulate an abundance of confirmations by our observing things that aren't *As*.[21] Clarifying this contretemps proves to be more difficult than one might expect; but some progress can be taken from comparing h's credence-space assays in relation to (a) this disjunctive evidence possibility vs. (b) the set of propositions that disjoins. Writing these with '$\sim x$' for '*not-x*,' '$x \vee y$' for '*x-or-y*,' and '$\cdot$' for '*and*,'

$$\text{(A1)} \qquad Cr(h) = \{Cr(h \cdot [d \vee e])\} + Cr(h \cdot \sim [d \vee e])$$

$$\text{(B1)} \qquad = \{Cr(h \cdot d \cdot e) + Cr(h \cdot d \cdot \sim e) + Cr(h \cdot \sim d \cdot e)\} + Cr(h \cdot \sim d \cdot \sim e)$$

$$\text{(A2)} \qquad Cr(\sim h) = \{Cr(\sim h \cdot [d \vee e])\} + Cr(\sim h \cdot \sim [d \vee e])$$

$$\text{(B2)} \qquad = \{Cr(\sim h \cdot d \cdot e) + Cr(\sim h \cdot d \cdot \sim e) + Cr(\sim h \cdot \sim d \cdot e)\} + Cr(\sim h \cdot \sim d \cdot \sim e)$$

[21] This is an updated version of Hempel's classic Ravens-paradox (see Google returns for "Hempel + ravens" and my widely ignored Rozeboom (1980) which makes a small point having large inference-theoretic import.)

wherein the curly-bracketed parts of (B1,2) are rationally equivalent to their curled counterparts in (A1,2).[22] When h entails d, the last two terms in (B1) are zero, but that no longer matters much here. What does matter is that the credence-space assay of *Cr(h)* in relation to $d \vee e$ shown in (A1,2), is logically equivalent to *Cr(h)*'s assay in relation to proposition pair $\langle d, e \rangle$ shown in (B1,2)—which seems to imply (correctly, I suggest) that whatever change in *Cr(h)* should be induced by change in $Cr(d \vee e)$ is identical to whatever response of $Cr(h)$ to an altered *Cr*-state of $\langle d, e \rangle$ is most rational.[23] What matters in this for confirmation of h is not whether its entailed disjunction's credence approaches certainty but what portions of that credence bear on h by what relevance connections obtain. Although how to determine the latter and readjust our *Cr(h)* accordingly very much remain open questions, this does deflate the threat of spurious confirmation by consequence dilution in hypothetico-deductive evidence appraisal and supports the thesis that when interpreting data as evidence for or against theories that interest us, we need to replace the hypothetico-deductive model of scientific inference as our operational gold-standard with some version of hypothetico-makeplausible.

References

Dale, A. I. (1991). *A history of inverse probability from Thomas Bayes to Karl Pearson*. New York: Springer-Verlag.

Harlow, L. L., Mulaik, S. A., & Steiger, H. H. (Eds.). (1997). *What if there were no significance tests*. New Jersey: Erlbaum.

Rozeboom, W. W. (1967). Why I know so much more than you do. *American Philosophical Quarterly*, *4*, 281–291.

Rozeboom, W. W. (1972a). 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. (1972b). Problems in the psycho-philosophy of knowledge. In J. R. Royce & W. W. Rozeboom (Eds.), *The psychology of knowing.* New York: Gordon & Breach.

Rozeboom, W. W. (1980). Nicod's criterion: Subtler than you think. *Philosophy of Science*, *47*, 638–643.

[22] Sketch of proof: For any propositions p, q, r, $(p \vee q)$ is logically equivalent to $(p \cdot q) \vee (p \cdot \sim q) \vee (\sim p \cdot \sim q)$ whose three disjunctive parts are mutually exclusive so that credence given to the whole should equal the sum of these parts isolate credences. And $r \cdot (p \vee q)$ is equivalent to $(r \cdot p) \vee (r \cdot q)$.

[23] This leaves open the possibility that whatever evidentiary input has induced change in the *Cr*-state of $\langle d, e \rangle$ may also affect *Cr(h)* through some cognate channel unmediated by its effect on $\langle d, e \rangle$ credence. It also raises the Pandora's-box question of how often or under what circumstances we predicate disjunctions that we don't realize are disjunctive.

Rozeboom, W. W. (1997). Good science is abductive, not hypothetico-deductive. In L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* New Jersey: Erlbaum.

Rozeboom, W. W. (2005). Meehl on metatheory. *Journal of Clinical Psychology*, *61*, 1317–1354.