To: David Lubinski
From: P.E. Meehl
Date: 1/5/99
Re: Thoughts on construct validity

[Don't take these as final thoughts, but because you are working on your big chapter for the *Annual Review*, I thought it useful to send along where I am in my thinking about this damnable problem. If I get clearer in the next few weeks or months, I may consider writing a paper, "Construct validity revisited," or "Construct validity after two generations," or "Three errors in Cronbach and Meehl," or some such. Lee has revisited it in various lectures, some of which has been published, and his views and mine may be so divergent by now that we couldn't coauthor such a paper. But I am setting aside publication questions for now, and am mainly trying to get clear in the head, with your help. Much of what I say here as my best current formulation is almost identical with Loevinger's classic paper (1957) and it would be pretty tiresome if I kept saying over and over, "… as Loevinger said …" Anybody familiar with her article will immediately see the content similarity. At this point, I think I am more in agreement with Jensen than with Humphreys about the philosophical aspects. I don't know about Carroll [1993], because pressures have deflected me from finishing his book, still on my agenda.

You know me well enough to realize that I view the history of science as a history of corrected mistakes. At my age and professional status, I am not so neurotic as to think I must defend everything I co-authored almost a half century ago on a difficult topic. On the other hand, I dislike being criticized for mistakes that we did not make. I am not going to bother mentioning various critiques of the CM55 [Cronbach & Meehl, 1955] paper, since the only one that shed much light was Loevinger's, and the others were sometimes obfuscating. They also illustrate Wesley Salmon's point that while scientists do not have to philosophize if they don't feel like it, they tend to do so when discussing theory, and they almost always do it badly.]

1. There are some infelicities of expression in CM55 and three major defects, more attributable to me than Cronbach. The first two are forcefully exposed by Loevinger (this is the *last* time I shall mention her—had our paper been co-authored by Cronbach, Loevinger, and Meehl, it would have been 50% improved). The third was

* After several years of correspondence and discussions, David Lubinski and Paul Meehl planned a comprehensive updated treatment of construct validity (working title, "Construct validity after 50 years"), targeted to coincide with the 50th anniversary of the Cronbach and Meehl (1955) citation classic, "Construct validity in psychological tests." Due to other commitments and priorities, then Meehl's death in early 2003, that article was not written. This memo to his friend shows Meehl's thinking about construct validity around the end of 1998. Minimal editing added references and explains a few shorthand notations for more general readership.—Leslie J. Yonce 9/8/2018

obliquely pointed at by Bechtoldt [1959], although due to his philosophical naïveté, he gets it wrong.

a. By listing four "kinds" of validity the way we did, we make it appear as if they were somehow of coordinate logical status, which they are not. Content, concurrent, and predictive validities constitute *the evidence for* construct validity, so it has a different status. I don't think a careful reader would fail to discern that epistemic relation from the complete text, where the discussion of internal consistency, factor analysis, external criterion keying, etc., as evidentiary paths to construct validation is pretty clear. But, nevertheless, making that *list* was misleading without an explicit statement about the different logical status of the four kinds.

b. Structural validity should certainly have been listed. Here again, it is puzzling why we didn't explicitly treat it that way, since we of course recognized that as one kind of evidence. One might suspect me as a Minnesota dustbowl empiricist, overly impressed with "blind external empirical keying" (MMPI, SVIB), to underestimate structural psychometric relations. But one can hardly imagine psychometrician and factor analyst Cronbach, author of the classic paper on coefficient α, to have ignored it. Here again, the text taken in entirety does not ignore it, so what we have is an unfortunate way of *summarizing* that doesn't square with the textual details.

   My guess is that we were skittish about saying anything that could possibly be taken to contradict the Technical Recommendations themselves, or clashed with committee members' opinions in discussion. That skittishness might have extended to the old "reliability doesn't prove validity" dictum, on which my recollection is the group had persisting disagreement. The internal psychometric structure of a test is traditionally considered a reliability matter. Although I have no memory of Lee and I discussing this in writing the article, it's hard to believe that we did not. Then in a few years Lee is writing the genralizability book, making them very close together. A mystery!

c. This is the bad one. In developing the notion of implicit definition of theoretical constructs via the nomological net, we never said *how much* of the net constituted the "definition." Do all of the lawlike relationships into which a construct like *g* enters (as a matter of empirical fact in the statistics!) constitute the implicit definition? Or, is it a proper subset of the nomologicals that defines the concept, the others to be treated as contingent statements that do not contribute to the meaning as such? I don't remember discussing this with Lee. The sole tape recording we have of one of our sessions when he visited the Philosophy of Science Center does not cover that point. It is a major weakness, because it permits Bechtoldt's attack that we have confused *meaning* with *significance*. I shall show below why that is a mistaken way of putting it, and I wish Embretson (1998) had not gone along with Bechtoldt on this matter. She certainly has no need to do this by way of defending her test construction methodology. He didn't get it right because he (*not* we!) still held simplistic operationism, which by that time had been abandoned for twenty years by all philosophers (with the exception of Bergmann, who was at Iowa and Spence's philosophical hatchet man). Bechtoldt had a point, but he didn't get it quite right.

I think I know the reason for *our* unclarity on this matter. I was at that time overly influenced by philosopher Wilfrid Sellars, who held that the whole law network defined each theoretical construct. He had written an almost unreadable article, "Concepts as involving laws and inconceivable without them" (1948), with which I believe Feigl, Scriven, and Maxwell disagreed. There is a subsequent discussion of that in the book that I have referred you to before (Feigl & Maxwell, 1961). Logician Carnap held that only a proper subset of the laws into which a theoretical entity enters constituted its implicit definition, and he accordingly referred to that subset as "meaning postulates." Whether this disagreement among the philosophers had become acute in the early '50s when we were writing CM55 I do not know, but I am sure it had surfaced. I, as the amateur philosopher, didn't have a definite opinion. Throughout our discussions of the paper, Lee was understandably trying to hold down the amount of technical philosophy to a minimum. In our only tape recording, Sellars is pushing the idea that if you're going to philosophize at all for the benefit of psychological readers, then you should do enough of it to get it right and make it clear. Cronbach is resisting manfully, reminding us about the unphilosophical audience. My hunch is that after our physical meeting, when we were completing the manuscript by correspondence, Lee would have thought it better not to get into that morass, especially since I didn't know where I stood. And I would have agreed with him. A purportedly clarifying paper which would be read by M.S. psychometricians and school teachers wouldn't help much by bringing up a technical issue in philosophy of science and then saying, "philosophers disagree about this, and we don't know what to think." Of course, that's intellectual history and not going to the merits of the issue.

[d. A fourth possible defect in CM55 would be our seemingly complete reliance upon the notion of implicit definition of theoretical terms. I believe it is accurate to say no contemporary logician has doubts about that view being *substantially* correct. However, within that broad framework, now universally accepted, one can raise two qualifier questions: (1) What about theoretical terms which are, in a sufficiently advanced state of knowledge, *explicitly* defined by means of other theoretical terms; and (2) some hold that there is an *interpretative text* (distinguishable from the "operational text" which ties the network directly to observations) that provides a further component of meaning. These are both important considerations, but they do not invalidate the main thesis. I can't discuss them adequately without some preliminary material, so I will relegate them to Sections 6 and 7 below.]

2. Operationism as a general philosophy of science, as an interpretation of how all theoretical concepts are defined, is dead, and has not been held by logicians for over 60 years. (That's an arbitrary timing, which I choose because 1936 was the date of Carnap's (1936-1937) classic paper on testability and meaning. A clearer date would be Carnap, Feigl, & Hempel writing in the 1950s.) It is, however, the case that some concepts are tied more intimately, more closely, more directly to operations-cum-observations than others. In that guarded sense, one can speak of a proper subset of theoretical concepts as being quasi-operationally defined (better, operationally quasi-defined). The most powerful theoretical concepts in the advanced sciences are typically removed by long derivation chains and numerous epistemic paths from what Eddington called "pointer-readings." But I have come to avoid even the weak operational quasi-definitional terminology, because when looked at closely, none of them turn out to be stipulative definitions. *They are conjectured but well-confirmed empirical linkages*. There is a simple metaphysical principle involved here, and it seems odd to me that psychologists have trouble comprehending it. I have been trying to explain this principle to Anne Anastasi for a half-century, and have given up. She can't seem to grasp the simple truth that if we ask, "What's the correlation between *x* and *u*?" (and the factor loading of observed

score x on factor u is their correlation), one *ipso facto* presupposes they are not always equal. Hence the two numbers cannot possibly numerify *the same entity*. The principle involved is Leibniz's Principle, the principle of the identity of indiscernibles. We have two expressions in a language and ask whether they denote the same individual. If everything that can be correctly said of one can be said of the other, then there's only one individual denoted. One way of defining the identity relation in symbolic logic is

$$(x = y) \equiv_{Df} (\forall P)\ (Px \leftrightarrow Py)$$

If every property or relation attributable to x is attributable to y and conversely, then x and y are the same individual. (If it seems odd to be talking first of x and y and then saying they're the same, it's like in ordinary algebra, where the notation using two variables to designate numbers does not commit us to saying that they can't be equal.) The predicates involved may be one-term predicates denoting properties, or they may be many-term predicates denoting relations.

Now, some of the properties that entities can have are properties differing in degree; that is, quantified predicates, so that if two expressions designate exactly the same entity, it would make no sense to ascribe different numerical values to the entities the expressions denote. Hence, it makes no sense to ask "how closely they are correlated," or "if x has a property, what is the probability that y has it," or "has it to the same degree." It follows from this that if there is an observational statement, whether qualitative or quantitative, that one speaks of as being *unreliable*, or *possibly biased*, *or capable of being improved upon* as a measurement procedure, then the observation cannot constitute the *meaning* of the measured property. That is, the attributed quality or degree is something *inferred from* the observation and, we hope, closely linked to the observation; but it is not literally the same as the observation. Thus, in physics, temperature is not literally defined as the height of a mercury column. Even before we had the kinetic theory of heat (which permits us to define temperature as the mean kinetic energy of the molecules) we already recognized that the nurse's reading of an oral temperature does not have perfect reliability, for a variety of reasons that everybody took for granted.

If someone thinks that a subject's electrodermal resistance is literally *defined by* the psychogalvanometer reading (that is, the arrow points on a scale of mho's or ohm's, to which the patient's arm is hooked up), Dave Lykken explains to him that this is a naive mistake on the part of the psychologist, who is neglecting the complexities of instrumentation. In his early work on the lie detector, Lykken showed it made a difference whether you had copper or silver electrodes, and how often and carefully they were cleaned. He showed that you should shave the patient's skin and sandpaper it lightly. (He had to dismiss a schizoid unperceptive research assistant, who kept sandpapering people until they became anxious and hence screwed up establishment of the proper baseline for the experiment.)

The Skinner box involves a simple response class, "operationally defined" by the fact that the rat gets the lever down, however he does it. But there is a distinction between the strength of a discriminated operant written in *Behavior of Organisms* as [$sS^D.R$] and the short-term response rate; that is, the momentary slope of the cumulative record. The rat draws pretty smooth curves, but if you want to mathematicize Skinner's [1938] system, you want to evaluate the analytic derivative at that point. That would mean deciding what kind of curve is fitted, and so on and so forth.

In every field of empirical science, we find that when we look closely at an "operational definition" of a concept, even before we get into any complex high-level theory, there is a difference between the reading that we get from the scientific instrument and the concept we are trying to "define." This being true, even for such simple unidimensional physical concepts as temperature and resistance and friction, and in the behavior area, lever-pressing response strength, it would be astonishing if it weren't true for such complicated matters as human traits. It holds there with a vengeance.

It's so obvious that one is embarrassed to even explain it. When I write the equation for a subject's test score in classical psychometrics, the last term is e_i, for *error*. Ergo—it follows as the night the day, for God's sake—the subject's observed score is not literally identifiable with any common factor, or composite of weighted factors (including the specific term s_i). A quantified something cannot be "defined by" an equation in a second something when for almost all instances (except when $e_i = 0$ for a very few individuals) their numerical values are different. Q.E.D., simple operationism is incorrect.

Consider something that at first glance seems almost as straightforward as lever-pressing strength, temperature, friction, or viscosity, say, *spelling ability*. It doesn't take much reflection to see how complicated things quickly become. Spelling ability ought to be easier to operationalize than social introversion or hypohedonia, but two complexities present themselves immediately. First, the domain of words to be spelled—which dictionary? There is evidence that if you want to estimate the percentage of terms that an individual *has occasion to spell* that he can spell correctly, inclusion of esoteric, high-level, low-frequency words in the sample may do your averaging procedure more harm than good. Such words will occur in any large random or stratified sample. If that were not so, we might treat the "operational" definition of spelling ability as the percent of words in a specified dictionary, say, the OED, that a person spells correctly. Many would challenge that choice of dictionary, but if we are super-conventionalists about operations, we simply *select* OED in order to *define*, come what may. It's a free country, huh? The second complexity is even worse, involving the format of the testing procedure. We can present a list of words with instructions to cross off the ones that are misspelled. Or, we can build a multiple-choice test where subjects must pick which is the correct spelling. Or, we can orally dictate words to be spelled. Or, we can sample the spelling of spontaneously produced documents. (A pragmatist would think that's the most realistic, "useful" way to do it.) Should that sample include only compositions produced in freshman English, or must we make sure to get access to the subject's personal letters and diary? Since most people don't keep a diary, would it be a distortion of our estimate of "average, representative dispositions to spell words correctly," if those who do keep diaries have their diaries included in our sample? None of the word domains sampled, and none of the formats of the spelling task, will give the same answer as any of the others, as is well-known. One might imagine some perfect Brunswik [1955] situational sampling. But I don't know how I would go about specifying the class of aspects of the spelling-demand environment which we would want to stratify and weight proportionally to the frequency with which one has to spell in these various contexts. The whole thing boggles the mind, which is why nobody has ever tried to carry out Brunswik's idealized program, even in the area of achievement testing.

But enough already. The point is obvious and, I believe, uncontroversial among psychometricians, whether they're in the personality, interest, achievement, or ability domain. The specification of an "operational definition" turns out to be rather one of postulating an *operational linkage* of some sort, a *factual* claim. It always involves a conjectured and partially confirmed network of correlations and causal relations. The *network* specification of a construct's *meaning* applies, even for the proper subset of psychometric concepts that we speak of loosely as being "operational."

3. Now comes the big question that you and I have been hassling over in our recent telephone conversations. How much of the network of relations to an inferred psychometric factor should be considered part of its *definition*? Well, that's not a question of fact or of mathematics, it's a meta-question. It's a matter of stipulation, the setting up of a convention. Here we must follow Popper's dictum that debating the meaning of words is the most useless of all philosophical enterprises. Even Nobel laureates can fall into this kind of morass, although not as badly as the ordinary language philosophers, whose entire enterprise is a waste of time. When Anderson found that cloud chamber path, having exactly the curvature that you get with a moving electron, and hence determining the same mass-to-charge ratio, except that it was bending the wrong way, he had discovered the positron, a particle predicted a couple of years earlier by Dirac in the equations which earned him the Nobel prize. Millikan, who was the electron big shot, wanted to call this new particle a 'positive electron,' and became quite vociferous about that verbal preference. Other people thought that phrase was clumsy, and they won the argument, which of course was not an argument either about facts or mathematics, but somebody's preference for a word. We don't want to get into that kind of useless semantic dispute.

The question is not about the *phrase* 'psychometric *g*,' but about a definitional policy, whatever expression is chosen for it—we could call it "general intelligence" if that were not politically incorrect—as to which are the strands of the postulated nomological network that we wish to include as part of the meaning of this node in the network. If some theoretically derivable relationship turned out to be empirically false, would we think we had to revise our conception of the entity?

[I'm going to say 'nomological' because it's the language of CM55, but if it weren't for the clumsy ugliness of it, I would insist on 'stochastological' (as I suggested in my Two Knights paper [Meehl, 1978]) because almost all of the lawlike relations we talk about in psychometrics are stochastic only, that is, not strict laws that we claim hold precisely except for measurement error. I haven't convinced the profession to use the term 'stochastological' but I still think it would be better.]

The first thing to keep crystal clear is that *g* is not a set of observable dispositions. *g* is an inferred hypothetical entity, a hypothetical construct in the MM48 sense [MacCorquodale & Meehl, 1948]. It underlies (produces, causes, influences) various testable dispositions in individuals, and we get at it via the phenomenon of individual differences in these dispositions. If people didn't differ in this respect, we would still have such a concept as intelligence, but we would not have a *g factor*, because the psychometric idea of a factor is inherently statistical and requires the phenomenon of individual differences. So, how do we get at it? What leads us to invent such a thing? Well, I am not going to go through the line of reasoning presented so ably by Jensen [1998], so suffice it to say that we begin with the notion of a cognitive task. Then we observe when we correlate people's quantitative dispositions in cognitive tasks, that we get positive manifold. That leads us to introduce this theoretical construct to explain these correlations. Whether we use ECTs* or the subtests of an omnibus IQ test, we seem to be tapping into something that shows up strongly, even in nonoverlapping task contexts, as in Bob Thorndike's [1987] classic study. So, psychometric *g* is something about the brains (yes, literally the brains) of people that influences their performance in a variety of tasks which may have almost nothing in common as regards manifest content (e.g., memory for digits, block design, and vocabulary) but which, if we try to characterize them, all seem to involve some aspect of information processing. There is no point in trying to sharpen a verbal definition of "information processing," as long as we can provide a wide list of examples and their factor loadings.

Then we find out that IQ, which turns out to be a pretty good estimator of *g*, even without optimal weighting in terms of factor loadings (presumably because of Wilks' theorem), turns out to be correlated with all sorts of other things that we observe in the life history of people rather than in the experimental laboratory or the psychological clinic. Table 9.3 in Jensen's book† lists impressive examples. We find that, while IQ is correlated .30–.40 with the SES of childhood family, it is correlated .50–.70 with one's achieved SES, but negligibly with one's childhood foster family's SES. This does not strike us as theoretically puzzling, especially in a society like ours.

But now suppose I study the incomes of stock brokers, and I get a negligible correlation with their IQs. I am surprised by this. I have a theoretical—and perhaps, if I am an I/O psychologist, a practical—reason for wanting to figure it out. Pending such an explanation, should this fact lead me to revise my *theoretical conception* of *g*? I say not. Why not? Well, the short answer is that it is many steps causally removed in the network from that property of the brain that I am getting at when I administer my omnibus IQ test or my ECT. *General principle*: The more causal

* [*Elementary cognitive task* (ECT) is a term used by Arthur Jensen (1998)]
† [Jensen, 1998, p. 300; reproduced from Brand, 1987.]

linkage steps there are between the theoretical concept and the fact that we have come to expect will be influenced, the lower the expectable statistical relation, for the simple reason that every link in the causal chain in the life sciences is almost certain to be stochastic rather than nomological because of the many other causal factors entering the chain at every linkage. The more probabilistic linkages there are between *g* and some remote criterion like stockbrokers' income, the more ways our predictions based on other studies of economic success may go awry. All of this was explained well, although without the mathematics, in a neglected paper by Doncaster G. Humm (1946).

4. Sticking to our implicit definition of *g* as the brain factor that influences omnibus IQ tests and ECTs very directly, we might proceed to investigate why stockbrokers' incomes fail to correlate with IQ when, in general, such income correlations are observed and make theoretical sense. It turns out to be easy to explain. The first thing we discover is studies in the literature (by economists and others) showing that stockbrokers' predictions of security prices are hardly better than chance, and that there are many individual stockbrokers who could just as well be flipping pennies. So, we have immediately two simple *statistical* explanations (causation aside) of the absence of correlation, the first being that Pearson's *r* depends upon range, and secondly, that an unreliable variable is not predictable by anything. We could stop there, because this explanation suffices. However, we might pursue it further, going into the details of individual predictions, broker by broker, applying measures of noncognitive attributes, such as cyclothymia, impulsiveness, average optimism/pessimism, laziness, 3-martini lunches, etc. (See Appendix II, FAX sent to you earlier.) We could perhaps show that while the task of a broker looks to be one of processing information, it only looks to be like that, and that the behavior output (advice to clients) has little or no correlation with the objective information. Or, we could look at another aspect: Suppose that there were some stable individual differences in information processing, although slight, how would they contribute to salary? It turns out that noncognitive attributes (Harvard accent, high-class pinstripe suits, greater age, distinguished silver hair) have a lot to do with salary. Commissions don't depend on successful predictions, they are a percent of the client's buying or selling transactions. There's a big batch of Paul Horst's "contingency factors," such as marrying the boss's daughter. Point: There is a mixture of purely statistical plus psychological plus sociological factors that enter into a stockbroker's income that, in the aggregate, have a heavy causal influence,

and since the stock market is not highly predictable by anybody, the information processing feature is only apparent, and plays little or no role in determining income. To the extent that we can measure these other nuisance factors and make appropriate corrections for them, along with discovering that the information processing component itself has small individual differences variance that's reliable, then even if we were not able to undertake all these microanalyses, we would not change our theoretical conception of *g* because of the “causal distance” stockbroker income is from that node in the net that we call *g*.

5. So, although Bechtoldt didn't say it right (I think, because he was hooked on logical positivism and operationism more than we were), when he makes the distinction between “meaning” and “significance,” claiming we conflated them, he is pointing to something defective in our presentation. He is wrong if he thinks that “meaning” can be provided without invoking *some* sort of network, as the spelling test example shows. But he is right in thinking that social impact or society's judgment or how much money you make, are not usefully considered part of the *meaning* of the term ‘intelligence.’ We didn't say it should be, but we didn't say it shouldn't, we left it unclear.

   I realize that remoteness in the net is a matter of degree, and I must think about whether it is desirable to say anything stronger about that than I have said. It may be that, having made that distinction, one should set up some sort of quasi-statistical criterion for what should enter into the definition. I don't know what to say about that. But it seems that I am closer to Jensen than to Humphreys, and I don't know about Carroll. The ideal situation would be this: Taking *g* as a node in the net, the only strands in the net that count as part of the implicit definition are those leading directly from the *g* node to some other node that is a direct behavioral disposition. It's obvious that income is not strictly a behavioral disposition. It is an indirect consequence of a whole class of variably correlated dispositions interacting with a bunch of environmental contingencies. Similarly, going to jail or having a car accident or being bankrupt, or getting syphilis, or any of the other things that are socially important and related to the bell curve, are not immediate *g*-produced behavioral dispositions. They are indirect and statistical *social impact and life course consequences*.

   If it should happen that some socially important average consequence of many individual behaviors had a large correlation with *g* but the behaviors involved in that long-term average social impact consequence did not involve processing information (scrutinized closely like an I/O psychologist doing a job analysis), we would not be happy with this. We would have an instance of the sort of thing Campbell and Fiske emphasized. We don't want something to be highly correlated with *g* if it does not involve, even indirectly, the processing of information; so the absence of a correlation of one sort, and the presence of a correlation of another sort, can both present us with *theoretical explanatory* problems. But they should not be allowed to alter the best interpretation we could get for *g* from what *g* influences in direct sampling of dispositions in the lab or in the clinic.

6. A problem arises for the contextual explication of theoretical terms. Some believe that the logician's technical device of the Ramsey sentence (see Appendix I) is not fully adequate as a metatheoretical explication. The theoretical terms get their meaning via their role in the net, the empirical component coming from the observational statements of the net. (I christened this the "upward seepage" theory of meaning in the early days of the Minnesota Center for Philosophy of Science.) It has been pointed out, and I agree, that the *embedding text* that gives empirical meaning to mathematical statements in the formalism of a scientific theory can be divided into two parts. One part is the *operational text*, by which the network is tied down to facts. It is what makes it a scientific theory rather than a "metaphysical" system. As Hempel (1952) puts it,

> A scientific theory might therefore be likened to a complex spatial network: Its terms are represented by the knots, while the threads connecting the latter correspond, in part, to the definitions and, in part, to the fundamental and derivative hypotheses included in the theory. The whole system floats, as it were, above the plane of observation and is anchored to it by rules of interpretation. These might be viewed as strings which are not part of the network but link certain points of the latter with specific places in the plane of observation. By virtue of those interpretive connections, the network can function as a scientific theory: From certain observational data, we may ascend, via an interpretive string, to some point in the theoretical network, thence proceed, via definitions and hypotheses, to other points, from which another interpretive string permits a descent to the plane of observation" (p.36).

But a different portion of the embedding text is not at the "operational" level, not directly connected with the observational evidence. It runs, so to speak, parallel with the formalism. What it does is *characterize* the theoretical entities and their connections—the nodes and the strands—higher up in the net, many steps removed from the data. It fills out their meaning over and above what is provided by upward seepage from observational statements. This *interpretive text* talks about the character of the theoretical entities. It typically uses a mixture of ordinary language with scientific and technical theoretical language.

An important class of terms in the interpretive text have the interesting property that they cut across completely different scientific disciplines. I have made a list of such words, some three dozen in number, raising the question whether they can be "Ramseyfied out" without loss of essential meanings. I don't assert they can't, but I have my doubts about that. Examples of these cross-disciplinary words are: 'accelerate,' 'adjoin,' 'annihilate,' 'approach,' 'be composed of,' 'combine,' 'conflict with,' 'countervail,' 'couple,' 'distance,' 'evolve,' 'extend,' 'facilitate,' 'fuse,' 'inhibit,' 'inner,' 'interact with,' 'interval,' 'link,' 'oppose,' 'persist,' 'potentiate,' 'prevent,' 'produce,' 'region,' 'resist,' 'retard,' 'separate,' 'subsequent to,' 'transform,' 'transmute,' 'turn into,' 'vanish,' 'vicinity of' (list from Meehl, 1990, *Corroboration and verisimilitude*…).

Note that, despite their being cross disciplinary—and in that sense, somewhat more "abstract" than terms like proton, libido, habit strength, cistron, velocity or money, that tell us what discipline we are talking about—they are not metalanguage terms. They are all in the object language. Furthermore, these generic terms play a crucial role at times in derivations that are not legitimated by the transformation rules of the formalism itself, a thing which philosophers of science have strangely failed to notice. (This can be fixed up but it's a bit technical and not useful here.)

Does the interpretive text, whether the terms are discipline-specific or generic, provide an exception to the view of implicit definition via the network? I think not. Whether the words used are scientific words or ordinary language words, all such words possess *their* meanings via a network although, in the case of terms from ordinary language, we are so accustomed to them, often from childhood on, that we are not aware of that. In psychometrics, the words are likely to be mentalistic terms from folk psychology. (There's nothing forbidden about folk psychology, it's just that we have to decide what portions of it are scientifically acceptable and which are not.) I believe it's universally agreed among ordinary language philosophers, as well as trait psychologists, that the terms of folk psychology involve a loose, primitive, preanalytic causal and structural theory of mind. So, the existence of an interpretive text as a subdivision of the embedding text that is not "operational" does not speak against the implicit definition view.

7. But isn't it the case that when a science becomes sufficiently advanced, and has a quasi-complete understanding of the composition and relations of the theoretical

entities, that we can sometimes offer *explicit* definitions of the theoretical terms? That is correct. But that still does not mean operational definitions in Bridgman's or Skinner's sense. It turns out that the other terms used in explicit definitions of our theoretical terms are themselves theoretical, not observational terms. Usually, but not always, the definer theoretical terms are those of a science at the next lower level in Comte's pyramid of the sciences. A good example is the gene. Prior to Watson, Crick, & Co., we already had a vast, ramified, and quite precise theory of the gene, very well confirmed: The concept of *gene* was defined implicitly by the network. Facts about linkage statistics together with the cytology of meiosis located the gene on the chromosome. These locations were then strongly corroborated by the statistics of aberrant pedigrees where direct cytological observation revealed displacements, translocations, deletions, and so on. But when I took genetics in 1940, they told us the gene was probably a protein, a notion that Crick was still entertaining seriously within a few weeks of Watson's structural breakthrough fiddling with cardboard models in the bar. I suppose that's why Watson's name comes first. But now we offer an explicit definition of the gene, namely, it's a cistron; and that is in turn defined as an ordered sequence of codons; and a codon is defined as an ordered triplet of adenine, guanine, cytosine, and thymine, together with a network statement about how each codon corresponds to one of the twenty amino acids, with some redundancy and a punctuation mark. So this looks like a case of quasi-complete knowledge about a theoretical entity that is no longer defined implicitly, but explicitly. That is true, in the sense that if the terms of biochemistry are taken as satisfactorily defined, then we have the kind of explicit definition, almost like Aristotle's definitions *per genus et differientiam* that we learned about in high school English or freshman logic. But these other terms are all *themselves* implicitly defined by the nomological network of biochemistry. So the explicit form of the reducibility does not refute the idea of implicit definition via the network, nor does it exemplify the metaconcept of operational definition that psychologists persist in requiring (but not really!).

8. In the case of psychometric *g*, the explicit definition would be in terms of a composite brain property. How do I know it's a brain property? Very simply; it's because you don't do cognitive tasks with your kidneys or bone marrow. The existence of a factor in the mathematical sense does not show, or even tend to show, that its physical composition must be of elements that are qualitatively homogeneous. I believe factor analysts have all accepted this ever since Brown's Thomson's argument about

his sampling theory. A nice analogy here is information in the military. Transmission of information, whether from the front to GHQ or orders in the other direction, plays a role in warfare that may be decisive. Some believe that Germany lost the First World War because Colonel Hentsch misunderstood, or badly conveyed, what the Supreme Command had in mind to General Alexander von Kluck in command of the First Army. When (if?) the North Vietnamese fired at the U.S. naval vessel in the Tonkin Bay fake episode, they were entitled under international law to do so because the ship was engaged in naval reconnaissance on behalf of the South, clearly classified as an act of war. If one factor analyzed a batch of battles using a rotation criterion in terms of who wins, or even without that, you would get a big fat information factor emerging from the statistics. It could be roughly quantified in some Shannon way as bits × distance ÷ time. In terms of winning the battle, that's all that counts. But the physical processes underlying this transmission of information are heterogenous, including human runner (Hitler), horseman, motorcyclist, carrier pigeon, heliograph, field telegraph, field telephone, field radio. The Russians lost the WW-I battle of Tannenberg partly because with incredible stupidity they were transmitting radio orders in clear. (Cf. Meehl, 1993.)

In considering the heritable component of *g*, one can imagine a half-dozen loci which, if independent, additive, and unlinked, would give us an expansion of the binomial $(\frac{1}{2} + \frac{1}{2})^{10}$ indistinguishable from a normal curve. I'm not suggesting only 10 genes, but it's interesting that 10 would be sufficient to give you that result. What could each of these genes control? Such microstructural features as number of cells (we have known for a century that mentally deficient subjects have fewer neurons in their brains), some microstructural property of distribution of terminal buttons over the synaptic scale, number and length of collaterals, number of neurons from which a neuron receives impulses, speed of transmission, rate of production and uptake of a neurotransmitter, and the like. Furthermore, Wilks' theorem tells us that the more loci are involved, the less important are their differential weights, so that if there were even 10 genes, we could almost neglect the weighting, except for very fine-tuned purposes, because the expected correlation between composites of two randomly assigned sets of weights will be better than $r = .90$.

9. Conclusion: Psychometric *g* is most conveniently defined by its role in the proximal region of the net. Its correlation with distal nodes, such as society's rewards to the individual, should be looked upon as contingent facts about the entity as “most

purely" defined. If some of these remote criteria surprise us, given our interpretive text (as in the broker example) this should usually be viewed as a problem of theoretical understanding and teasing apart the obfuscating variables rather than as an invitation to amend the definition.

Appendix I: The Ramsey Sentence

David Lykken, in his contribution to the Meehl *Festschrift*, says I am the only person who can fool people into thinking the Ramsey sentence is interesting. My point in doing that in our seminar on methodology is that whether or not it's interesting, it's important. There aren't very many things in symbolic logic that can be of use to the psychologist. I never lean on the students to learn the technicalities of it, about which I am no expert myself, despite my interest in philosophy. But the Ramsey Sentence is important because it is a more rigorous explication of the idea of *implicit definition* than the metaphor of a "nomological network." For students who don't like arguments from analogy and don't feel comfortable talking about the network, I tell them the Ramsey Sentence is the logician's formal development of the idea. The reason the idea is important is that many psychologists, raised in simplistic operationism or the earlier forms of logical empiricism, think that one and the same expression or system of expressions cannot both *define* and *assert* concurrently. They are right for single sentences, but not for related sets of sentences. Since there is no scientific theory, even the most impoverished, that consists of only a single sentence, the traditional point is not important in understanding empirical science. Some criticisms of Cronbach and Meehl (1955) (e.g., Bechtoldt, Brodbeck, I think Astin) were based partly on that error. Of course, if I define the word 'crow' as "a large black bird that caws and eats carrion," then the proposition "all crows are black" can't be an empirical claim, refutable by observations, because being black is part of my definition. This we learned in high school English, and again in freshman logic. But when we consider a theory consisting of at least two sentences, we get a qualitative change. Now the semantic overlap of the sentences may be such that an undefined ("theoretical") term appears in each of the sentences but has dropped out in the conclusion, so that even a simple Aristotelian syllogism in Barbara can illustrate it as a degenerate, pitiful, two-sentence theory, thus: "All large black birds that caw are corvinal." "All corvinal birds eat carrion." Ergo, "All large black birds that caw eat carrion." For somebody who does not know Latin, the term 'corvinal' has no connotation, so he doesn't have a clue what it might mean. But he needn't know what it means to draw a valid conclusion from these two premises. Since the word 'corvinal' has disappeared in the syllogistic inference (being Aristotle's *middle term* M that mediates between S (*subject*) and P (*predicate*) of the conclusion) he now has derived a statement in the observation language. Consequently he can test the theory by getting the observational facts.

The logician would say that he also has, in addition to a testable empirical theory, a *partial interpretation* of the mysterious term ‘corvinal.’ It turns out—I would say, it's intuitively obvious—that in empirical science, even the most advanced empirical sciences, all of the theoretical terms are, at best, partially interpreted. We would never know that we had a complete interpretation of a term like ‘electron’ unless we knew everything that could be known about electrons. And we have to *know* that we know everything that could be known, which obviously, we cannot know. This led Feyerabend to say in one of our Center discussions, “Meehl, you talk about Pap's open concepts; well, all concepts are open,” to which Imre Lakatos replied, “Yes, but then, some concepts are opener than others.”

So I tell the students in our methodology seminar that we won't snow them with a lot of symbolic logic, but we will use a bit of symbolic logic in explaining the Ramsey Sentence, and they should courageously put up with it. If they really understand the Ramsey Sentence, and the idea of implicit theoretical definition via the network, they will understand science better than the majority of psychologists.

So here's about the Ramsey Sentence: If I do decide to expand this memo into an article on revisiting construct validity, I am going to fight with the editor to include an explanation of the Ramsey Sentence. We expect psychologists to learn calculus and matrix algebra if it helps them theorize or appraise theories, and we expect them to develop some handiness with electronics if they're in the lab, and to learn some biology if they're interested in heritability of traits. It seems strange to me that we don't expect them to learn those parts of logic, including elementary symbolic logic, that have relevance to thinking about scientific theories.

The Ramsey Sentence is named for Frank Plumpton Ramsey, a student of Bertrand Russell's and the mentor of Richard Braithwaite (whose book, *Scientific Explanation* (~~1950~~ 1953) is still very much worth reading). Ramsey died in his late 20s, I believe of peritonitis from appendicitis. He was a mix of logician, mathematician, and economist, and did not publish much during his short lifetime. After his death, Braithwaite found among Ramsey's papers a number of manuscripts. Braithwaite edited them, some of which appear in a posthumous collection called *Foundations of Mathematics and Other Logical Essays*, still in print [Ramsey, 1931]. One of the essays was on theories, and was simply called “Theories.” The Ramsey Sentence idea is buried in a short cryptic paragraph of that essay. Ramsey was here motivated by Bertrand Russell's parsimonious and error-phobic maxim, "Whenever possible, substitute logical constructions for inferred entities." But he deviated from Russell in wishing to preserve the inference to entities we don't observe, the "surplus meaning" of Reichenbach (1938) and MacCorquodale and Meehl (1948). Ramsey conceives of the theory to have been axiomatized. We write down all of the postulates, their conjunction is the theory as usually stated. Some of them are "pure," composed of only the theoretical terms plus the formalism of logic and mathematics. Others are "mixed," containing at least one observational and one theoretical term. The conjunction of pure theoretical postulates is T; the conjunction of mixed postulates is C (for "coordinating," by Carnap). The theory as initially expressed is TC, and it uses the theory-terms τ_1, τ_2, … τ_k (e.g., electron, gene, libido, habit strength, g factor, enzyme). From conjunctions of suitable subsets of these postulates, one derives theorems from which, as in the crow syllogism, *the theoretical terms have disappeared*. Now we have an observation statement and the derivability of such is what makes the theory an empirical theory.

We now start reading through the conjunction TC of the postulates, and as soon as we hit a theoretical term, τ_1, we erase it and replace it with a dummy variable, say θ_1. Continuing to read through the theory, every time we come across the same theory word, we substitute the θ_1 variable, which has no antecedent connotation. We then bind θ_1 by the existential operator ("there exists") in the front. Then we read through again and find a different theoretical word, τ_2, erase it, and substitute, in every locus in which it appears, the variable θ_2, and then we bind θ_2 with the existential operator. We continue in this way until we have erased all of the theoretical words and replaced them by dummy variables θ_1, θ_2, … θ_k. The variables' subscripts have no metrical or ordinal meaning, but are simply what the statistician calls "football numbers," that is, they merely serve to keep straight where each theoretical word recurs in the *structure*, i.e., how is it related to the other theoretical and observational terms by the postulated theoretical laws. When we have completed this so there aren't any theory words left, they having all been replaced by θs, and the θs having been bound by the existential operator, we have the Ramsey Sentence of the theory, written ^{R}TC. Instead of the original expression in Ts and observation Os,

$$TC = F[\tau_1, \tau_2, \ldots \tau_k, O_1, O_2, \ldots O_j]$$

we now write

$$^{R}TC = (\exists\theta_1)\ (\exists\theta_2)\ \ldots\ (\exists\theta_k)\ [\mathrm{F}(\theta_1, \theta_2, \ldots\ \theta_k)\ O1, O_2, \ldots\ O_j]$$

where the only descriptors are observational Os. The "structural relations" among the concepts has been preserved, but the theory-terms have been eliminated. The theory is said to have been "Ramseyfied," and the theoretical terms, which have now disappeared are said to have been "Ramseyfied out." The only nonformal terms that are left are observational terms occurring in the observational theorems derived at the extreme right. Ramsey's point in this logician trickery was to show how you could *eliminate the theoretical terms without eliminating the theory*.

What is the Ramsey Sentence saying? It says there is a θ_1, and there is a θ_2, and a θ_3 such that [and now follows some kind of relationship among the three θs that is specified by the mathematics and logic of the formalism]. So, if somebody wants to know what the libido is, we tell them, "Well, the libido is something (a whatsit or whoozis, or whatever) that does this and that and the other, and is influenced by such-and-such, and changes in the following manner, and so on, and so on." That's how the physicist explains to you what an electron is. Nobody asks the physicist what an electron really is in itself, or in its essence, which sounds like you wanted to know how it feels to be an electron, or what does an electron, so to speak, experience from the inside. There are some technical logician problems about the Ramsey Sentence which are, I believe, still under discussion. But despite some needed cleaning-up jobs, it has a very important role, as we see Carnap gave to it in his last book (*Foundations of physics: An introduction to the philosophy of science*. Martin Gardner (ed.), 1966).

It troubled him ("for years," he says) that the Ramsey Sentence, combining *assertion of fact* with *explication of meanings*, seems at first like a breakdown of the distinction between analytic and synthetic propositions, and hence could give ammunition to logician Quine who opposes the sharp distinction. Carnap's trick (nothing wrong with a logician's trick to solve a logician's puzzle) is to write a third sentence that relates the Ramsey Sentence to the original theory sentence. Let TC be the theory sentence containing the theory terms $\tau_1, \tau_2, \ldots\ \tau_k$, and ^{R}TC be the Ramsey Sentence containing only the dummy variables $\theta_1, \theta_2, \ldots\ \theta_k$. Then he writes a new sentence A_{T}

$$A_{\mathrm{T}} : {}^{R}T \supset T$$

which is analytic. A_T asserts that *if* the Ramsey Sentence is true, *then* the (original) theory sentence is true, given the correct interpretation of the τs. All observational facts derivable from ^{R}TC are derivable from TC, and conversely. TC and ^{R}TC are both synthetic, they both say something about the world, and both could be false. But A_T is analytic, it asserts nothing about the world but only a logical relation between the other two. It is called the "Carnap Sentence" of the theory.

Appendix II: Memorandum PEM → DL, 12/19/98

To : D. Lubinski
From : Paul E. Meehl
Date : 12/19/98 (faxed)
Re : Implicit definition in construct validity (Big Memo in process)

The attached diagram shows why I prefer (a *volitional bifurcation*, Reichenbach 1938) to treat only psychometrics and ECTs‡ as comprising the implicit definition of g. Bechtoldt had a point but didn't get it right.

Let $x_1, x_2, \ldots x_n$, be dispositions directly and immediately observable in the subject's behavior (psychometric subtests, ECTs).

Let $y_1, y_2, \ldots y_n$, be subject's behavior dispositions in a variety of non-psychometric, non-ECT contexts (e.g., learning speed, judgment).

Let $z_1, z_2, \ldots z_n$, be responses of other persons to their sample of the subject's ys.

Let $v_1, v_2, \ldots v_n$, be long-term consequences of the zs (e.g., income, jail, status).

Then we have causal-and-statistical chains originating in g, links are sets of xs, ys, zs

$$g \to \{x_i\} \to \{y_i\} \to \{z_i\} \to \{v_i\}$$

Each link in chains is stochastic rather than nomological, because other variables (not listed, they're intrusive “side-chains”) exert causal influence. The more remote, the worse.

So we implicitly define g by the *local* network of xs, the direct link, $g \to \{x_i\}$.

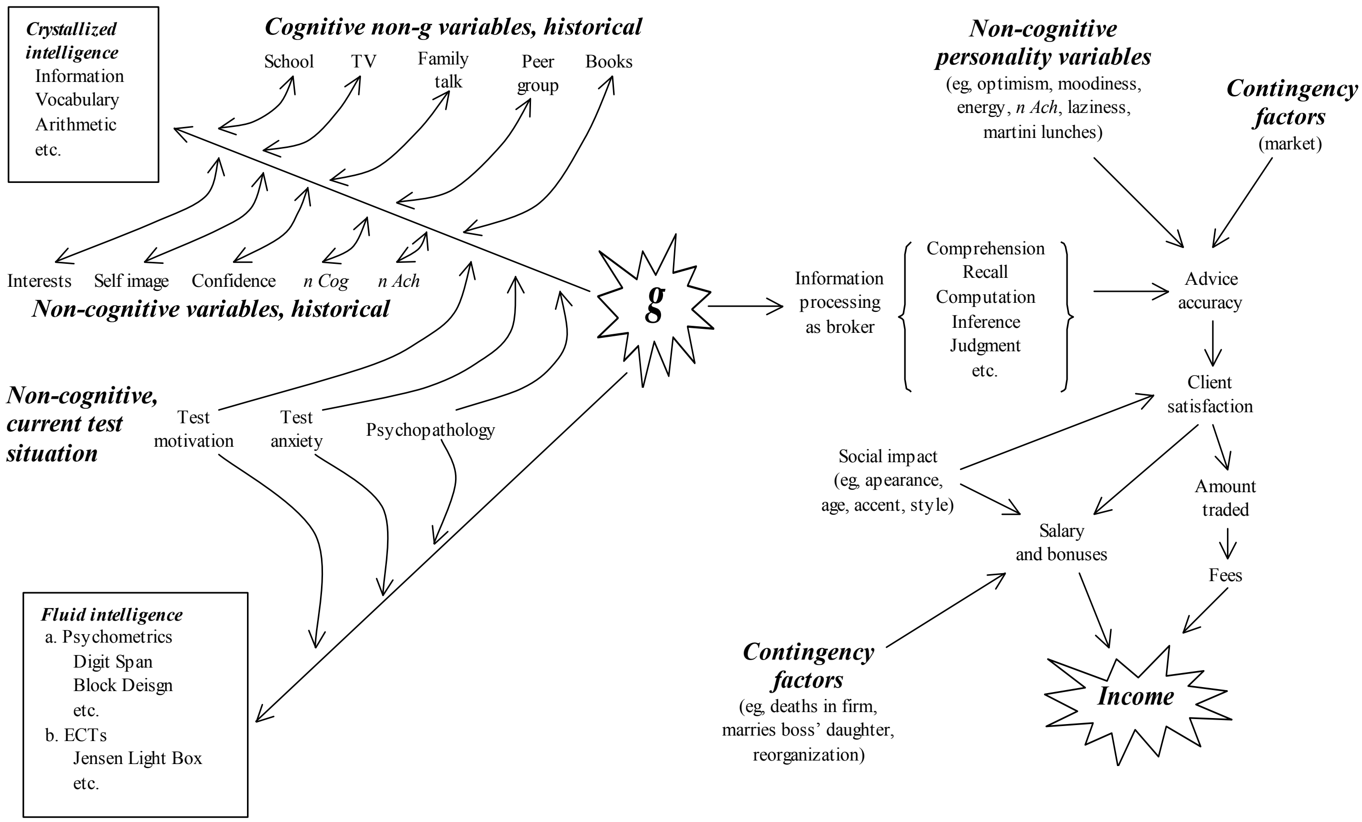
[If we have objective ys (e.g., measured—not *rated*—work sample) we may include them. We almost never do, so I don't include them. A rating on “smarts” is not a y but a z, not included!]

So in network terms, no linkage involving a mediating node—two sequenced strands—is included in the implicit definition.

[Intrusive causes weakening the strands between g and xs are not mediating nodes, rather they are arrows sticking into path diagram. We minimize them or hold them quasi-constant (e.g., school exposure, test motivation) but can't eliminate them, which is one reason why even the best g-loaded tasks don't correlate perfectly with g.]

---

‡ ECTs, elementary cognitive tasks (Jensen, 1998).


Crystallized intelligence
Information
Vocabulary
Arithmetic
etc.
Cognitive non-g variables, historical
School
TV
Family talk
Peer group
Books
Interests
Self image
Confidence
n Cog
n Ach
Non-cognitive variables, historical
g
Non-cognitive, current test situation
Test motivation
Test anxiety
Psychopathology
Fluid intelligence
a. Psychometrics
Digit Span
Block Deisgn
etc.
b. ECTs
Jensen Light Box
etc.
Information processing as broker
Comprehension
Recall
Computation
Inference
Judgment
etc.
Non-cognitive personality variables
(eg, optimism, moodiness, energy, n Ach, laziness, martini lunches)
Contingency factors
(market)
Advice accuracy
Client satisfaction
Social impact (eg, apearance, age, accent, style)
Amount traded
Salary and bonuses
Fees
Contingency factors
(eg, deaths in firm, marries boss' daughter, reorganization)
Income

References

[Bechtoldt, H. P. (1959). Construct validity: A critique. *American Psychologist*, *14*, 619-629.]

[Braithwaite, R. B. (1953). *Scientific explanation: A study of the function of theory, probability and law in science*. New York: Harper.]

[Brand, C. (1987). The importance of general intelligence. In S. Modgil & C. Modgil (Eds.), *Arthur Jensen: Consensus and controversy* (pp. 278-283). New York: Falmer.]

[Brunswik, E. (1955) Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*, 193-217.]

Carnap, R. (1936–37). Testability and meaning. *Philosophy of Science*, *3*, 420-471; *4*, 2-40. Reprinted with corrigenda and additional bibliography, New Haven, CT: Yale University Graduate Philosophy Club, 1950. Reprinted in H. Feigl & M. Broadbeck (Eds.), *Readings in the philosophy of science* (pp. 47-92). New York: Appleton-Century-Crofts, 1953.

Carnap, R. (1966). *Philosophical foundations of physics*. New York: Basic Books.

[Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, UK: Cambridge University Press.]

[Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–302.]

Embretson, S. E. (1998). A cognitive design system approach to generating valid tests: Application to abstract reasoning. *Psychological Methods*, *3*, 380-396.

Feigl, H., & Maxwell, G. (Eds.). (1961). *Current issues in the philosophy of science*. New York: Holt, Rinehart and Winston.

Hempel, C. G. (1952). *Fundamentals of concept formation in empirical science*. *International encyclopedia of unified science*, II, no. 7. Chicago: University of Chicago Press.

Humm, D. (1946). Validation by remote criteria. *Journal of Applied Psychology*, *30*, 333-339.

[Jensen, A. R. (1998) *The g factor: The science of mental ability*. Westport, CT: Praeger.]

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, *3*, 635-694.

MacCorquodale, K., & Meehl, P. E. (1948). On a distinction between hypothetical constructs and intervening variables. *Psychological Review*, *55*, 95-107.

[Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, *46*, 806-834.]

Meehl, P. E. (1990). *Corroboration and verisimilitude: Against Lakatos' "sheer leap of faith"* (Working Paper, MCPS–90–01). Minneapolis: University of Minnesota, Center for Philosophy of Science.

Meehl, P. E. (1993). Four queries about factor reality. *History and Philosophy of Psychology Bulletin*, *5*(No. 2), 4-5.

Ramsey, F. R. (1931). *Foundations of Mathematics and Other Logical Essays*. Paterson, NJ: Littlefield, Adams.

Reichenbach, H. (1938). *Experience and prediction*. Chicago: University of Chicago Press.

Sellars, W. (1948). Concepts as involving laws and inconceivable without them. *Philosophy of Science*, *15*, 287-315.

Skinner, B. F. (1938). *The behavior of organisms: an experimental analysis*. New York: Appleton-Century.

[Thorndike, R. L. (1987). Stability of factor loadings. *Personality and Individual Differences*, *8*, 585-586.]