COMMENT

The Tragedy of Operationalism

Mark H. Bickhard
LEHIGH UNIVERSITY

ABSTRACT. Operational definitions were a neo-Machian development that connected with the positivism of logical positivism. Logical positivism failed, with the failure of operational definitions being just one of multiple and multifarious failures of logical positivism more broadly. Operationalism, however, has continued to seduce psychology more than half a century after it was repudiated by philosophers of science, including the very logical positivists who had first taken it seriously. It carries with it a presupposed metaphysics that is false in virtually all of its particulars, and thereby distorts and obscures genuine issues concerning the nature of theory and of science. It makes it particularly difficult for psychologists, under the thrall of this dogma, to free themselves from these false presuppositions, and to think about, create and critique genuine scientific theory and process. That is the tragedy of operationalism.

KEY WORDS: logical positivism, operational definition, philosophy of science, positivism

One of the central assumptions of logical positivism was that all meaning is empirical. Empirical content, and only empirical content, constituted genuine cognitive significance. A strong motivation for this assumption was the power it presumably provided for banishing metaphysics: metaphysical sentences had no empirical content, and therefore were simply meaningless. An early development of this radical empiricism was the attempt to construct a verifiability theory of meaning: the meaning of a sentence is its means of verification—the means by which its truth value could be determined (Suppe, 1977).

An earlier version of positivism, Machian neo-positivism, held that science was solely in the business of discovering patterns in empirical data. Science is a kind of cultural eye that 'sees' such vast patterns, and all legitimate theory is 'just' notational abbreviations for such patterns. Such a view has grave difficulties accounting for the truth values of sentences—how can a data pattern *per se* be true or false?—and foundered on the strongly meaningful uses of mathematics in relativity theory and quantum mechanics. There was no way to construe, for example, Minkowski space-time as a data pattern.

THEORY & PSYCHOLOGY Copyright © 2001 SAGE Publications. VOL. 11(1): 35–44
[0959-3543(200102)11:1;35–44;015767]

Previously, mathematics was often dismissed as a means of calculation, therefore mathematical sentences were not empirically meaningful—they were just instrumental rules for such calculation—and, therefore, mathematics did not violate the strictures on all meaning being empirical. Logical positivism developed, among other reasons, as an attempted correction of this difficulty for the raw empiricism of neo-positivism: relativity and quantum mechanics used mathematics in ways that could not be dismissed as mere calculational devices. Minkowski space-time, to continue that example, is not only not a data pattern, it is also not a mere calculational device. In logical positivism, logic and mathematics were construed as tautologies, by grammatical convention, and therefore had no empirical, no cognitive, significance, even though they were legitimate sentence forms—including in relativity theory and quantum mechanics.

The verifiability theory of meaning was an attempt to account for sentence meaning. It carried forward the Fregean point that sentences, not words, are the fundamental unit of meaning. The fundamental pole for meaning was sentential truth value. Bridgman's proposal for operational definitions focused on sub-sentential meanings, word meanings generally. In this, it hearkened back to Machian positivism more than to logical positivism.

The philosophy of science that pervades psychology has borrowed from both logical positivism and neo-positivism, but the Machian view of science as a discerner of data patterns, and of all meaning as being constituted in such patterns, is still the core of the scientific mythologies that permeate psychology (Bickhard, 1992; Smith, 1986). There are numerous errors that accompany this mythology, but that of the ideology of operational definitions is among the most pernicious. Operationalism carries with it many of the additional errors of positivism as background assumptions, making them that much more difficult to recognize and transcend.

Logical positivism failed for many reasons. The failures of the verifiability theory of meaning and its successors were among them, but the failures went far beyond this inability to make good on the claim that all meaning is strictly empirically constituted. Operational definitions are a kind of pragmatic version of the logical positivist bridge laws, laws that presumably connected theoretical terms with data. But the entire edifice collapsed, with operational definitions *per se* as just one of the side-shows that had been rejected long before the overall project was acknowledged—even by its proponents—to have failed.

Not so in psychology, however. For multiple reasons, the Machian view of science that underlies operational definitionalism has remained the dominant conception of science in psychology. This is very slowly decreasing, but positivism still dominates the central institutions and ideologies of the field. This is in spite of the half a century lag from when the philosophy of science repudiated logical positivism, and the even longer lag from when operational definitions were given any serious philosophical credence.

Operational definitionalism holds that *meaning* is constituted in *empirical operations*. It identifies the processes of *measurement* or identification, and of *testing*. The empirical methods for measurement constitute meaning; and a scientific test is constituted as the carrying out of such an operationally defined measurement. This multiple conflation among meaning, measurement and testing distorts conceptions of science and procedures of science. It is deeply pernicious.[1]

The target article, 'On the Failure of Operationism' (Grace, 2001), illustrates such distortions. I will address these and related problems mostly in the order in which they appear. 'It might seem tempting, based on the failure of operationism, to reject a positivist approach to psychology outright (e.g. Bickhard, 1992; Leahey, 1983)' (p. 7). This stands the relationship on its head. The multiple and multifarious failures of positivism, *including* those of operationalism, provide very good reasons for rejecting operationalism *per se*.[2]

'. . . Bridgman advocated a strict, positivistic empiricism: the elimination, insofar as possible, of metaphysics and a priori principles . . .' (p. 8). The impossibility of so eliminating metaphysics from science was one of the powerful reasons why logical positivism failed, and it lies behind many of the reasons why operationalism is so pernicious: the radical empiricism of operationalism makes it difficult to understand how science does, in fact, involve theoretical and metaphysical assumptions, and must involve them, and thereby makes it difficult to think about and to critique those assumptions. Because this assumption of the desirable eliminability of metaphysics is itself a metaphysical assumption, we find here a self-protective neurosis of scientific mythology.

'He [Bridgman] agreed with Mach that the task of science was fundamentally *empirical description*' (p. 9). This is a bald statement of one of the pernicious pieces of baggage connected to operationalism that was mentioned above. The author neither bothers to make the claim that this is in fact a correct or desirable conception of science—a claim which would contradict the last fifty years of the philosophy of science, and the last hundred years if the differences between logical positivism and Machian neopositivism are taken into account—nor does he attempt to show that operationalism does not, or need not, pull along with it this deeply damaging assumption.

In a discussion of whether the Bohr model of the atom implies that space at that atomic level is Euclidean, Grace comments: 'For that assertion to be demonstrated [that atomic scale space is Euclidean], it would be necessary to show that any model that made non-Euclidean assumptions would make different (and incorrect) predictions than Bohr's model' (p. 10). Note that here we do not find a discussion of the operational definition of a concept, but, rather, the testing of a prediction. Such testing would certainly involve various measurements, but none of them would constitute the meaning—the operational definition—of 'Euclidean geometry'.

'Bridgman has therefore proposed a criterion by which we may ascribe ontological status to an unobservable construct: evidence for such a construct must be obtained in at least two sets of independent operations' (p. 10). Such evidence, as above for Euclidean geometry, will presumably involve measurements. But, again, those measurements will not constitute the meaning of the construct, and may well be measurements of phenomena quite different from that of the construct *per se*, so long as the theory involving the construct makes predictions involving those measured variables. Here we have a straightforward conflation between testing and measurement.

'[I]t proved very difficult to provide an operational definition for dispositional concepts' (p. 11). Yes, it did, and none was ever found. As mentioned, if sugar being soluble means that 'if put into water, it will dissolve', then *anything* not put into water will, by this definition, **not** be soluble.

In response to this and other problems, Bridgman retracted the claim that 'a concept was synonymous with its corresponding set of operations. Rather, an operational definition was a *necessary*, but not *sufficient*, constituent for scientific meaning' (p. 11). The logical positivists made similar moves attempting to salvage their empiricism in the face of such problems, all to no avail. They also recognized that even the move to a partial empirical rendering of dispositional terms—for example, that *X* being soluble implies that, if put into water, then *X* will dissolve—already constituted an abandoning of the strict empiricism of the program. Solubility, and virtually all other scientific concepts, would have meanings not fully capturable by such empiricist means (Suppe, 1977). Solubility could not be merely a data pattern; science could not be merely a matter of description: Mach was wrong.[3]

I find the discussion concerning operations as measuring vs as producing phenomena to be of interest, but of indirect relevance to the deepest problems with operationalism. So I will not comment on this discussion *per se*.

'In this passage, Bridgman is emphasizing (in his own language) the importance for measurement of reliability and validity . . .' (p. 19). Reliability and validity are of fundamental importance for measurement, but they do not constitute meaning. In an earlier comment, we find just this conflation: 'Stevens' real error was in assuming that an unobservable construct was scientifically valid when defined with only a single set of operations' (p. 16). First, is validity a property of a concept, or of one or more measurements *relative to* a concept? The conflations involved here make this a pointless question: the concept is *defined* by its measurement procedures. Second, defining a concept by one set, or one hundred sets, of operations does nothing to solve the problems of, for example, capturing dispositional concepts in strictly empiricist terms.

'Here Skinner is questioning the logical positivists' correspondence theory of meaning, namely that a meaning can be found to correspond to every sentence . . .' (p. 20). This is *not* the correspondence theory of meaning. The correspondence theory of meaning holds that meaning is constituted in correspondences between the logical structure of sentences (not necessarily the surface structure) and facts or purported facts in the world. Wittgenstein's *Tractatus* is a classic—and classically influential on logical positivism—version of a correspondence model. The logical positivists held that all meaning was empirical—in effect, a matter of correspondence—but that some sentences were acceptable even though of no cognitive (empirical) significance because they were statements of tautologies based on grammatical conventions—e.g. logic and mathematics. There was never an assumption that there was a meaning for every sentence. In fact, ridding the world of metaphysics with the claim that the sentences of metaphysics were meaningless was a major appeal that logical positivism offered. But even logic and mathematics were held to have no meaning. So, not having a meaning was not equated to being in the nether-world of metaphysics. Most certainly, however, no assumption was ever made by the logical positivists that 'a meaning can be found to correspond to every sentence'.

'All research involves making philosophical assumptions that cannot be justified purely on either rational or empirical grounds' (p. 22). Well, certainly not purely on *empirical* grounds. With that partial caveat, 'Yes, precisely', but what then happens to the 'everything is descriptions of empirical data patterns' metaphysics of operationalism? And if the metaphysics of operationalism is bankrupt, why isn't operationalism bankrupt? Why would anyone wish to defend operationalism, which is based precisely on the assumption that all meaning is empirical, and at the same time acknowledge that meaning goes beyond all possible empirical constitution?

'The importance of the method of converging operations is that it provides the only means of making potentially valid inferences about unobservable constructs or processes' (p. 24). This is simply false. It presupposes precisely the conflation among meaning, testing and measuring that permeates the entire article. Suppose, within an information-processing framework of background assumptions, a hypothesis is developed that two tasks differ by one extra processing step in one of the tasks relative to the other task. One way of testing that might be to look for the extra time delay that would be presumably created by that extra step. For reasons of reliability, we might even find two or more ways to measure that delay, or we might also postulate an increase of error rate because of the additional step during which error could be made. So, we have: (1) a model in which the meanings are constituted in presumed information processing terms; (2) a prediction based on the model of an extra step involved in one task relative to another; (3) more specified predications of increased reaction time and perhaps error based on this extra step of processing; and (4) one or more ways of measuring reaction time and error.

Do the reaction time and error measurement procedures 'define' that information-processing step? That is an absurd notion, yet operationalism would require that they do. Do the measurement procedures for reaction time and error *converge* on an extra step of, say, performing an extra transformation on an underlying syntactic form? Again, absurd. The measurement procedures are being used to test one or more predictions that are derived from an underlying model whose meanings are constituted in information-processing terms, not in measurement terms. Meaning, hypothesis testing and measurement procedures are all involved here, but they are not identical. The conflations of operationalism make such distinctions technically impossible, and therefore practically difficult to make and to understand and to teach. Again, the perniciousness, the tragedy, of operationalism.

These conflations are again demonstrated in a note at this point concerning quarks:

> Although the quark model has become standard, a single quark has never been directly observed, in spite of numerous attempts to do so and the fact that as particles with fractional electrical charges they ought to be easy to identify .... The question of whether quarks are 'real', that is, whether they are tiny bits of matter and not just a model that describes the symmetries in 'real' particles, remains unresolved. (p. 28, n. 2)

This is utterly false. The quark model was originally accepted as involving real quarks on the basis of testing hypotheses about constituents of protons and neutrons. This testing involved high-energy scattering experiments. Evidence consistent with such parts of protons and neutrons was in fact obtained. But no one would claim that the scattering results or procedures *defined* quarks. A great deal of further work has continued to support the quantum chromo-dynamic theories involving quark fields.

Early in the research regarding quarks, much effort was expended in attempting to find single, isolated quarks, but none was ever found. Contemporary theory, however, predicts precisely that result. In contemporary theory, the energy required to separate an isolated quark is also sufficient energy to create a pair of additional quarks from the quantum vacuum. Those additional quarks then pair up with the quarks that might otherwise become isolated, so we find (minimally) pairs of quarks instead of single quarks.

This theory might well be wrong. There are some clear ways in which it is consensually agreed that it *is* wrong: for example, it cannot incorporate gravity in its current form. But it will never be disproven, it will never fail an empirical test, by virtue of failing to find an isolated quark—that failure is precisely one of the predictions of the theory (Kaku, 1993; 't Hooft, 1997; Weinberg, 1996).[4] Yet it is a very well tested theory, and the existence of quarks is tentatively accepted because such existence is part of the theory, and because the theory has such powerful overall support, not because some operational definition of quarks has been proposed. The very idea is simply silly.

Quark theory involves so many kinds of potential predictions that to presume to define quarks in terms of the measurement procedures involved in testing those predictions is, again, simply silly. Quarks are not scattering patterns; quarks are not patterns of particle decay and interaction; quarks are not particle 'jets' emerging from high-energy collisions; quarks are not hidden dark matter in the universe; and so on. Yet all of these are involved in various tests of the basic theory—tests that might involve, for example, measuring the velocities of stars rotating within a galaxy and computing how much mass must be present in that galaxy to account for such velocities; noting that there is not sufficient mass in visible light; postulating various potential sources of such hidden mass; checking various predictions of quantum chromo-dynamics and its family of theories for their predictions about kinds of particles as yet unconfirmed; and noting whether or not it is possible to account for the hidden mass in a way that is consistent with the overall theory of which quarks are a part. Not one single piece of these complex steps defines the concept 'quark'. Neither singly nor collectively do they constitute an operationalization of quarks. Nor taken together with all the other myriads of ways in which quark theory can be and has been tested, and all the even greater number of measurement procedures involved in all of those actual and potential tests.

Quark theory is a massive refutation of operationalism. But, then, so is special relativity, and general relativity, and quantum electro-dynamics, and information-processing theories, and any other real science.

String theory is another refutation of operationalism from contemporary physics. String theory offers the possibility of integrating quantum field theory with gravity—something that has thus far eluded all efforts. For this reason and others, string theory has been a focus of excitement in physics since the early 1980s. But the mathematics of string theory is so difficult—and of a new kind—that decades of work have yet to derive empirically testable consequences of the theory. String theory could easily turn out to be wrong; it is certain that it will be rejected if no testable consequences are found. But it is currently a domain of a great deal of work by some of the best minds alive (Greene, 1999; Kaku, 1988; Witten, 1996, 2000).

Yet, the size of strings, according to current theory, is so small that directly detecting a string *qua* string would require so much energy as to be unimaginable. There is no direct observation of a string that is physically possible. Of even greater consequence for operationalism, however, is that string theory has yet to make contact with empirical testing. This is a deficiency that all string theorists hope to remedy. But the important point for current purposes is that this lack of empirical grounding does not render work on string theory irrational, nor does it render the terms of string theory devoid of meaning. Operationalism forces the view that theory grows up out of empirical data—there is no other way that the terms of a theory could be meaningful if those terms didn't constitute stand-ins for patterns of data. In this view, string theory is impossible. It is *irrational* since it doesn't conform to 'correct' practice of developing theory out of data, but it is *impossible* because the terms of string theory could not have any scientific meaning at all—because they do not make contact with empirical data, and they most certainly do not have, nor is it possible for them to have, operational *definitions*.[5]

This corruption of understanding the nature and power of theory that is forced by operationalism is yet another of the tragedies of operationalism. It contributes to the lack of understanding of theory in psychology and to the relative naïveté of the theoretical work that does exist in psychology.

'[R]ather, the network as a whole was to be evaluated with respect to its predictions regarding observables' (p. 24). Setting aside that this nomological network model too did not last, it raises relevant questions on its own: Is the *testing* of *predictions* (of a nomological net) the same as operationalizing a concept? Is the *testing* of a nomological net the same as operationally defining unobservable concepts in that net? Again, operationalism forces 'yes' answers in both cases, and is simply wrong in both cases.

In a discussion of Campbell's rejection of operationalism: 'But the dogma that he [Campbell] rejects is psychologists' naïve operationism of the 1930s, which maintained that an unobservable variable was defined by specifying a single set of operations used to measure or produce it' (p. 28, n. 3). No. Campbell's rejection was of the dogma that meaning could be constituted in operations—any number of operations, measuring or producing or whatever.[6]

'[A] *convergent* operationism provides the only potentially valid method of making inferences regarding unobservable constructs on the basis of empirical data' (p. 26). It is of interest that the issue has shifted here from defining concepts to making inferences about them. What kinds of inferences? Made in what way? And so on. An obfuscating rewording of the basic issues is being used here. In fact, even with this rewording, the statement is still false. Inferences regarding unobservable constructs are made, and 'validly' made, all the time regarding quarks, information-processing steps and procedures, and so on—and on—on the basis of tests of the hypotheses derived from the theories involved. No *operationism*, no operational *definitions*, need be involved, convergent or otherwise.

Measuring is important; being precise about measurement procedures is important; reliability and validity are important; testing hypotheses and theories is important; scientific meaning is important; theories are important. But none of them can be equated to any of the others. Yet operationalism conflates all of them, thus muddying our thinking about these issues. That is the tragedy of operationalism.

Notes

1. In important respects, this conflation is and was a regression even from 19th-century notions of science. Electrical and magnetic fields, for example, could not be directly observed. Instead, theories involving them were to be tested via predictions derived from them (Laudan, 1981). Such distinctions were lost in the Machian view that psychology adopted—inspired by, among other influences, that of the dogma of operational definitions.
2. I will not attempt to speak for Leahey, but it is simply false that Bickhard (1992) made any such claim. The argument, rather, was, as in the annotated sentence in the main text, that operationalism must be rejected because of its own failures *and* because it pulls along multiple other errors of positivism. That is, the failure of positivism more generally is a (good, additional) reason for rejecting operationalism, not the other way around.
3. Grace's version of the changes Bridgman made in the quoted sentence above is not quite the change that the logical positivists made. It is also not clear what 'constituent' means in this context. But these points need not be pursued here.
4. There is also, of course, the possibility that it will be demonstrated that that prediction of 'no isolated quarks' cannot in fact be derived from, or made consistent with, the underlying theory, in which case the absence of isolated quarks will again be an anomaly. But, even in this scenario, the anomaly will not be because of an absence of operational definition, or the failure of operational definition. (How could a definition fail, anyway? That is just one more manifestation of the illicit conflation between measurement and testing.)
5. If and when string theory does become empirically testable, it will not be via operational definitions. It will be in terms of testable consequences about particle decays, or testable consequences for cosmology, or some similar set of highly indirect consequences. Once again, operationalism forces a confusion between testing hypotheses derived from theories and operationalizing terms in the theory. And then it invites a confusion between operationalizing terms in a theory and observing entities to which those terms refer—witness the claim in the target paper that testing the reality of quarks depends on operationalizing, observing, single quarks. After all, if the meaning of a theoretical term is just the data pattern to which it refers, then any term in the theory is *defined* in terms of its special data pattern, and observation of an instance of that pattern *is* observation of an instance of what the term refers to.
6. Without getting bogged down in textual analysis, I will simply report that Don Campbell (personal communication) was very pleased with the points made in Bickhard (1992). In case there is any residual question, the model of representation that is adumbrated in Bickhard and Campbell (2000) is strongly anti-empiricist.

References

Bickhard, M.H. (1992). Myths of science: Misconceptions of science in contemporary psychology. *Theory & Psychology*, *2*(3), 321–337.

Bickhard, M.H., & Campbell, D.T. (2000). *Variations in variation and selection: The ubiquity of the variation-and-selective-retention ratchet in emergent organizational complexity*. Unpublished typescript, Lehigh University, Bethlehem, PA.

Grace, R.C. (2001). On the failure of operationism. *Theory & Psychology*, *11*(1), 5–33.

Greene, B. (1999). *The elegant universe*. New York: Norton.

Kaku, M. (1988). *Introduction to superstrings*. New York: Springer.

Kaku, M. (1993). *Quantum field theory*. Oxford: Oxford University Press.

Laudan, L. (1981). *Science and hypothesis*. Dordrecht: Reidel.

Leahey, T.H. (1983). Operationism and ideology: Reply to Kendler. *The Journal of Mind and Behavior*, *4*, 81–90.

Smith, L.D. (1986). *Behaviorism and logical positivism*. Stanford, CA: Stanford University Press.

Suppe, F. (1977). The search for philosophic understanding of scientific theories. In F. Suppe (Ed.), *The structure of scientific theories* (2nd ed.; pp. 3–241). Urbana: University of Illinois Press.

't Hooft, G. (1997). *In search of the ultimate building blocks*. Cambridge: Cambridge University Press.

Weinberg, S. (1996). *The quantum theory of fields: Vol. II. Modern applications*. Cambridge: Cambridge University Press.

Witten, E. (1996). Reflections on the fate of space-time. *Physics Today*, *49*(4), 24–30.

Witten, E. (2000). Magic, mystery, and matrix. In V. Arnold, M. Atiyah, P. Lax, & B. Mazur (Eds.), *Mathematics: Frontiers and perspectives* (pp. 343–352). Providence, RI: American Mathematical Society.

MARK H. BICKHARD is the Henry R. Luce Professor of Cognitive Robotics and the Philosophy of Knowledge at Lehigh University. He is interested in the development of process models of psychological phenomena, with a particular focus on the nature of representation and related issues in cognition. Publications include: 'Interaction and Representation', *Theory & Psychology*, 1999, *9*(4), 435–458; 'Levels of Representationality', *Journal of Experimental and Theoretical Artificial Intelligence*, 1998, *10*(2), 179–215; 'Troubles with Computationalism', in W.O'Donohue and R.F. Kitchener (Eds.), *The Philosophy of Psychology* (Sage, 1996); and (with L. Terveen) *Foundational Issues in Artificial Intelligence and Cognitive Science: Impasse and Solution* (Elsevier Scientific, 1995). ADDRESS: Department of Psychology, 17 Memorial Drive East, Lehigh University, Bethlehem, PA 18015, USA. [email: mark.bickhard@lehigh.edu]