Pergamon

Person. individ. Diff. Vol. 21, No. 1, pp. 115–118, 1996

0191-8869/96 $15.00+0.00

S0191-8869(96)00061-X

WHEREOF ONE CANNOT SPEAK, THEREOF ONE MUST REMAIN SILENT

Jeremy S. H. Jackson and Michael D. Maraun

Department of Psychology, Simon Fraser University, Burnaby, B.C., Canada VSA 1S6

(Received 8 February 1996)

When an individual claims that "these are measurements of σ", or "the way to measure the 'σ' of 'δ' is...", his claims are not made correct or incorrect via the products of empirical investigation. Regardless of whether σ is a physical or psychological concept, the justification for such claims comes from: (1) the existence of a rule guided practice for measuring σ; and (2) having correctly followed the rules that govern the practice. Rules (standards of correctness) are a *precondition* to distinguishing between claims that are correct and claims that are incorrect. Why is this the case? A measurement claim has the form: the σ of δ is x, in which x is standardly given in particular units. The practice of measuring the σ of δ in particular units is the basis for a *grammatical* relation between σ and δ, for it rests on the meanings of both δ and σ, and meaning is manifest in the rules for the correct use of concepts (Wittgenstein, 1953; Ter Hark, 1990). The justification for the statement that voltmeters measure voltage, not electricity, is given in comparison to rules that establish the meaning of electricity and voltage. Similarly, the rejection of the sensation seeking scale (SSS) as a measure of sensation seeking follows from careful scrutiny of the grounds for the instantiation of a disposition, in contrast to a need, preference, attitude or desire (see for example: Alston, 1975; Wittgenstein, 1953; Ter Hark, 1990; Rozeboom, 1984; Tuomela, 1978). Let us briefly review relevant theory.

(1) A rule cannot be discovered, established via empirical facts or detected from a careful study of the empirical

In a given context, the only thing that could establish the claim that "there is a rule ϕ" as correct or incorrect is a comparison to the rule itself (if there is a rule in play at all). It is obvious, however, that such a comparison *presupposes* an understanding of the rule itself. Hence, there is no such thing as empirical evidence establishing the nature or existence of a rule. Rules are autonomous and non-discoverable (Ter Hark, 1990).

(2) Measurement practices, tied as they are to systems of rules (i.e. standards of correctness), are logically bound to exhibit autonomy with respect to empirical cases and investigations

No empirical finding can refute or support a measurement claim. For example, the claim that "*these* are measurements of IQ" cannot be shown to be correct or incorrect on the basis of actual numbers recorded, nor the correlation of these numbers with other sets of numbers (e.g. measurements of school performance). On the contrary, rules are *constitutive* for empirical evidence: *these* empirical findings are not about IQ at all unless they *already* are based on numbers that have meaning as measurements of IQ. The correlation between the sets of numbers x and y is not the correlation between, e.g. IQ and school performance, unless x and y are already justifiable as measurements of IQ and school performance. Instead, the autonomous correctness of the measurements would necessitate the explanation of some possibly idiosyncratic facts about, for example, measurements of IQ. One might counter that if a set of numbers were presented as "measurements of IQ," and included the number 500, this would be an example of how empirical *facts* (i.e. that humans cannot have an IQ of 500) can legislate on the felicity of measurement claims. However, to know that humans do not have IQs of 500 *presupposes* a practice of measuring the IQ of humans, and this practice is grounded in rules for measuring. Knowledge about the IQ of humans is accumulated by measuring correctly (i.e. by following the rules). If challenged now to justify the

claim that "humans do not have IQs of 500," one would ultimately have to cite the rules for the correct measurement of IQ. In other words, what makes this a fact is that it too was generated according to correct measurement practices. Thus, while empirical facts can be used to cast doubt on the accuracy of particular measurements, they themselves are dependent on the system of rules that provide *justification* for measurement claims, and thus are no more foundational than the claim that is in question.

Construct validation is a peculiarity in the history of science. It arose out of a perverse mix of positivism and empirical realism (Norris, 1983) that paved the way for a fatal conflation of the meanings of psychological concepts (established by grammatical rules) and the empirical results, facts and theories they denote. In paragraph 10, when Dr Zuckerman states that "We are not of the school of *meaning* which insists 'if the data don't fit the *theory* the data must go' ", he unwittingly slips into this standard construct validity conflation of theory and meaning. It is precisely this conflation of empirical and conceptual issues that has resulted in an endemic failure to grasp the autonomy of the constitutive rule systems that are the foundation of meaning and measurement. The problem is that in construct validation theory, *knowing* about something is confused with an understanding of the *meaning* of the concept that denotes that something: "Scientifically speaking, to 'make clear what something *is*' means to set forth the laws in which it occurs" (Cronback & Meehl, 1955, p. 290). This is mistaken. One may know more or less about *it*, build a correct or incorrect case about *it*, articulate to a greater or lesser extent the laws into which *it* enters, discover much, or little, about *it*, only as long as there exist rules for the application of the concept that denotes *it*. Furthermore, one must be prepared to cite these rules as justification for the claim that *these* empirical facts are about *it*. One cannot even begin the exercise of coherent empirical investigation, hypothesis construction, theory building, etc., without an understanding of the meaning (i.e. grounds for application) of the concept that will organize and denote this work. Contra Dr Zuckerman (paragraph 5), meaning is not a private matter with test constructors and respondents each having differential access. This view echoes the private language position that was summarily routed by Wittgenstein in the mid 1900s. Although it is well beyond the scope of this rebuttal to review the salient points, we might ask Dr Zuckerman rhetorically, how could people communicate? We *teach* the meanings of concepts to our children, *correct* one anothers misuses, get into arguments and seek the favour of authoritative sources, etc. In short, *meaning* is public and shared. What is personal and idiosyncratic are a person's experiences, feelings, opinions, etc., which (incidentally) are dependent upon language for their expression [see Lakoff (1987) for a recent discussion].

The consequence of construct validation's conflation of meaning/measurement with empirical discovery is a plethora of category errors and striking absurdities. Consider the following two paradoxes induced by CVs conflation of meaning and discovery.

(1) Discovery/validation conflation

Consider the construct validation tenet that a tests validity rests on its conforming to certain empirical conditions. Assume that for test T to be a valid measure of θ, T must have a large positive correlation with variable X and that initially $r(T,X)$ is, in fact, large and positive. Since T is now seen to be a reasonable measure of θ it is employed in research projects designed to accumulate empirical facts about θ. Over the course of a year, however, it is noticed that $r(T,X)$ moves toward zero. Now, is this an important new discovery about the relationship between θ and X, or evidence that T is no longer a valid measure of θ? Since, according to construct validation theory, the result implies Ts invalidity, it is not possible to make an empirical discovery about the stability of the relationship between θ and X using T. Since any result is potentially a part of the nomological net of θ and T, this discovery/validity ambiguity applies to *all* empirical results involving T. Moreover, it is not merely a case of uncertainty as to whether a given result is, in fact, about validity or discovery, for in construct validation theory there exists no criterion to distinguish the two. Construct validation bars the usual criteria, the rules of language, for distinguishing between empirical discovery and meaning and hence perversely renders them the same thing.

(2) Nomological paradox

Consider the construct validation tenet that, through progressive approximations, empirical discoveries can clarify the meaning of a concept. Assume that discoveries about θ are made with

measurement instrument T. Now, for T to provide relevant empirical findings, i.e. findings about θ, it must be in line with the meaning of θ (metaphorically, it must have been created in the image of θ). Assume that it is found that $r(T,Y) = 0$ and that we agree that this result changes our understanding of the meaning of θ. This implies that originally we did not have a correct conception of the meaning of θ. Since T was constructed in accord with this earlier understanding of the meaning of θ, T is now seen to be an improper measure of θ. It follows then that $r(T,Y) = 0$ was not really about θ—it was not denoted by θ. Hence, the paradox is that the 'result' $r(T,Y) = 0$ undermines itself: The fact that $r(T,Y) = 0$ means that this result has no meaning! Zuckerman unwittingly provides a prime illustration of this absurdity with his series of SS 'definitions' (each constituted of a mix of hypothesis, theory, correlate and stipulation) that are laid down without any bridging rules to link them. Such behaviour is decidedly not 'in the spirit' of major traditions in science. For example, in an attempt to reformulate Newton's circular definition of mass, Mach (1960) follows a long recognized necessity in which the introduction of a new definition is accompanied by logical support and rules that link the competing formulations.

We now close by bringing this analysis to bear on some specific statements made in Dr Zuckerman's commentary.

- In paragraph 5, Dr Zuckerman calls to the authority of other test constructors on the fundamental issue of what correlation says about meaning: "Jackson and Maraun say that item intercorrelations have nothing to do with the meaning of items. Test constructors know this is not so. The more similar the meaning (to the test constructor and presumably the respondent) the higher the correlation between the items." Let us examine this claim in more detail. In the first place, whether items that are 'similar conceptually' also correlate is a case by case empirical issue. What is required to address this issue in a particular context is, of course, a criterion for 'conceptual similarity' (see, e.g. Guttman, 1977; Ter Hark, 1990; Lakoff, 1987). But what is essential to construct validity is the converse: items that correlate *must* be 'similar' conceptually. And this requirement manifestly does not hold. The illustration of this confusion was precisely the reason for the introduction of the traffic exposure example in which highly correlated items were, in fact, conceptually unrelated.
- Contra Dr Zuckerman (paragraph 2), length and breadth are not two distinct *dimensions*, but two distinct *concepts*. Area is a *concept* that is defined as a function of length and breadth: $A = L \times B$. One has a measurement of area if one has correctly measured L and B, and has employed this formula correctly.
- In paragraph 2, we are told that "It is true that correlations between such dimensions do not imply their identity but they may define a higher order dimension." This is nothing but a conflation of the *meanings* of two concepts a and b with the geometric concept of dimension inherent to the structure of the data they denote (perhaps in a common factor space).
- Dr Zuckerman's comments on the place of criterion, content and construct validity in this discussion require comment. For the record, criterion validity refers to nothing more than applied prediction and so bears not on the question of the validity of a measurement claim. However, the shabby comparison of content validity with our discussion of the *grammar* of measurement is unforgivable. To speak of domain sampling in the first place *presupposes* a criterion of admissibility to the domain (i.e. rules that specify what counts as an element of the domain). This, for example, is why Guttman (1971) developed facet analysis.
- Finally, we must respond directly to Dr Zuckerman's unsupported allegation that we draw inappropriate analogies between psychological and physical measures. While there are obvious differences, they are both rule-guided practices. In the same way that there are rules for the instantiation of the concept of height, for example, there are also rules for the instantiation of psychological concepts. What are the rules that Dr Zuckerman should have paid attention to? The rules for the instantiation of *dispositional* concepts that are discussed at length by, to name a few,

Tuomela (1978), Rozeboom (1984) and Wittgenstein (1953). As we have shown, these rules exclude items that instantiate needs, attitudes, preferences and desires as candidates for a scale purporting to measure the *dispositional* concept of sensation seeking. Since, in the development of the SSS, Dr Zuckerman pays no attention to the grammatical rules for the instantiation of dispositional concepts, he must ultimately remain mute on issues pertaining to the concept of sensation seeking.

The problems of psychology are some of the most difficult and worthwhile in science. But a much greater sophistication is needed if we are to have the productive, respected science we all desire. As Norris (1983, p. 53) states in discussing construct validation theory, "there is a lack of penetrating criticism and of alternative views in discussions of the field." This dogmatic allegiance to Cronbach and Meehl's archaic treatment of measurement will not do. We must make getting it right, rather than getting it published, the guiding principle of psychological investigation.

REFERENCES

Alston, W. (1975). Traits, consistency and conceptual alternatives for personality theory. *Journal of Theory of Social Behaviour, 5*, 17–48.

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

Guttman, L. (1971). Measurement as structural theory. *Psychometrika, 36*, 329–348.

Guttman, L. (1977). What is not what in theory construction. In Borg, I. (Ed.) *Multidimensional data representation: When and why*. Ann Arbour, MT: M.A. Thesis.

Lakoff, G. (1987). *Women, fire, and dangerous things*. Chicago, IL: University of Chicago Press.

Mach, E. (1960). *Science of mechanics. A critical and historical account of its development*. Lasalle, IL: Open Court.

Norris, S. (1983). The inconsistencies at the foundation of construct validation theory. In House, E. (Ed.) *Philosophy of education*. San Francisco: Jossey–Bass.

Rozeboom, W. (1984). Dispositions do explain: Picking up the pieces after hurricane Walter. In Royce, J. & Mos, L. (Eds) *Annals of theoretical psychology*, Vol. 1.

Ter Hark, M. (1990). *Beyond the inner and the outer*. Dordrecht, The Netherlands: Kluwer academic publishers.

Tuomela, R. (1978). *Dispositions*. Dordrecht, The Netherlands: Reidel.

Wittgenstein, L. (1953). *Philosophical investigations*. Oxford: Basil Blackwell Ltd.