Theory
& Psychology

Measurement Theory, Psychology and the Revolution That Cannot Happen

Günter Trendler
UNIVERSITY OF HEIDELBERG

ABSTRACT. Doubt is raised that revolutions in measurement theory, for example conjoint measurement or Rasch measurement, will lead to the quantification of psychological attributes. First, the meaning of measurement is explained. Relying on this, it is demonstrated that in order to attain quantification under causally complex circumstances it is necessary to manipulate the phenomena involved and control systematic disturbances. The construction of experimental apparatus is necessary to accomplish these tasks. The creation of modern quantitative science through the adoption of this method is called the Galilean revolution. Next the Millean quantity objection is formulated. If the Galilean revolution is not possible in psychology, the task of quantification is not solvable. The objection is defended. Psychological phenomena are neither manipulable nor controllable to the required extent. Therefore they are not measurable.

KEY WORDS: laws of nature, measurement theory, quantification, quantity objection

> *... it is true that if there were no phenomena which are independent of all but a manageably small set of conditions, physics would be impossible.*
> (Wigner, 1960, p. 4)

Recently, Joel Michell (2000, 2004) has described quantitative psychology, and psychometrics in particular, as a pathological science. He points out that psychometricians never seriously attempt to find out if psychological attributes are really measurable, but assume that they are. Similar objections have been raised before (Brown, 1934; Fischer, 1968, 1974; Wright, 1997; Wright & Linacre, 1989). But Michell goes a step further. According to him, psychometrics additionally suffers from a methodological thought disorder. Not only has the quantitative hypothesis never been satisfactorily tested, but this fact is purposely disguised by a flawed conception of measurement.

THEORY & PSYCHOLOGY VOL. 19 (5): 579–599

DOI: 10.1177/0959354309341926 http://tap.sagepub.com

However, Michell (1990, 1997, 1999) believes that the situation isn't hopeless. Psychologists must simply start testing the hypothesis that psychological attributes are quantitative. He praises the theory of conjoint measurement as developed by Luce and Tukey (1964) as a revolution in measurement theory, because it may finally allow the measurement of psychological attributes. Rasch models for measurement (Rasch, 1960/1980) are celebrated for similar reasons (Andrich, 1988, 2004; Blais, 2003). Rasch models and probabilistic models in general arouse much more interest among psychologists because they seem to account for the non-deterministic behavior of human beings. It is argued that in psychology deterministic models for measurement are inappropriate since 'what a human being actually does seems quite haphazard, none less than radioactive emission' (Rasch, 1960/1980, p. 11). Therefore supporters of probabilistic theories prefer, like in modern physics, 'models which are *indeterministic*, where *chance* plays a decisive role' (Rasch, 1960/1980, p. 11).

In my view this confidence is not well founded. As I will argue, it is not a revolution in measurement theory that psychology needs, but rather what I will call the Galilean revolution. As Kurt Lewin (1931) emphasized, it is the transition from the Aristotelian to the Galilean scientific method which must take place in psychology in order to transform it into a quantitative science. Unfortunately Lewin missed some important aspects of Galileo's innovative experimental method. Before I expand upon this assertion I will clarify what measurement is, why we can measure some attributes, and how we can find out if they are measurable. Next I will describe the role of the Galilean revolution for the task of quantification. Finally I will explain why I think that it is very unlikely that the latter will ever take place in psychology.

Measurement and Quantitative Structure

As Michell (1997) discovered in a survey of the literature, Stevens' definition of measurement (1951) still enjoys a widespread acceptance among psychologists. This is unfortunate since, as Michell explains, it is extremely misleading. Firstly, it leads psychologists to believe that measurement consists entirely of 'the assignment of numerals to objects or events according to rules' (p. 1). An immediate unfavorable result of this definition is that 'anything can be conceived as measurement, for conventionally one can always find or stipulate some rule, with the help of which numerals could be assigned to diversified aspects of things' (Berka, 1983, p. 24). Secondly, it ignores the fact that the question of measurability entails an empirical issue, namely the question of whether the attribute involved is really quantitative. To avoid misunderstandings it is therefore important to make clear the meaning by which measurement should be understood.

First of all, it must be mentioned that representational measurement theory is at present the dominant measurement theory (Krantz, Luce, Suppes, &

Tversky, 1971; Luce & Suppes, 2002; Narens, 1985; Narens & Luce, 1986; Pfanzagl, 1968; Scott & Suppes, 1958; Suppes & Zinnes, 1963). It is grounded in the general theory of models as was developed by Tarski (1954). The most basic concept of this theory is the concept of a relational system or structure. A relational system consists of a finite set of elements, called the domain of the relational system, and relations between these elements. It is a purely formal system and as such contains abstract symbols. This means that in order to apply model theory, for example, to mathematics, the symbols must be interpreted. The elements of the relational system may be mathematical entities (e.g., numbers) in which case the relations between the entities are mathematical relations (e.g., equality, order relations, etc.) or operations (e.g., addition, multiplication, etc.). Such a relational system is called a numerical relational system. If it contains empirical objects (e.g., a set of rods of different length), the corresponding system is called an empirical relational system. Another important concept of model theory is that of homomorphism between relational systems. According to representational measurement theory, measurement can be defined as being a homomorphism between empirical and numerical structures.

According to 'the classical view', however, measurement always involves the concept of quantity (Helmholtz, 1887/1998; Hölder, 1901, 1924). This concept of measurement has recently been revived by Michell (1990, 1997, 1999). What are quantities? A quantity (e.g., length, mass, temperature, etc.) is a kind of property empirical objects can possess which admits variation in terms of magnitudes (Ellis, 1966; Michell, 1990). Magnitudes are specific levels of a quantity (e.g., the length of an object is a magnitude of the quantity 'length'). If we interpret a relational system in such a way that the elements are magnitudes of a quantity (e.g., a set of rods of different length) and that the relations between elements are relations between magnitudes (i.e., relations of equality, order and additivity), then we obtain a quantitative structure. As Hölder (1901) demonstrated, a homomorphism between ratios of magnitudes of a quantity and the system of positive real numbers exists only if the relations between magnitudes satisfy certain so-called 'conditions of quantity' (or axioms, as they are usually called in representational measurement theory). For example, such conditions are: any two magnitudes of the same quantity are either identical or different (e.g., two rods are either equal in length or not); for every magnitude of a quantity there is another that is less (e.g., for any rod there can be found or manufactured another one that is shorter); for every pair of magnitudes there is another that is less (e.g., if two rods are concatenated, there can be found or manufactured another one which is shorter); and so forth (for a complete exposition of the conditions of quantity see Michell, 1999, chap. 3).

Classical measurement theory provides some important advantages over representational theory. For instance, in contrast to the representational definition, the classical definition of measurement specifies more strict selection criteria for what is constitutive as an empirical relational system. It should therefore come as no surprise that representationalism allows empirical interpretations of relational structures which, according to the classical view, certainly would not be regarded as measurement, since they do not fit the concept of quantity (for an example, see Krantz et al., 1971, pp. 87–88). Furthermore, as Michell (1997) points out, the classical concept is closer to measurement as it is defined in the physical sciences, namely as '*the estimation or discovery of the ratio of some magnitude of a quantitative attribute to a unit of the same attribute*' (p. 358).

In conclusion, *an attribute must satisfy the conditions of quantity in order to be measurable*. Hence, if we want to know if psychological attributes are quantitative in the same sense as physical quantities, we must ask if they possess quantitative structure. This is in my view the only non-trivial question to be asked with regard to the measurability of psychological attributes. Notice that, in this sense, no psychological attribute has ever been measured (Blinkhorn, 1997; Cliff, 1992; Schönemann, 1994).

Conditions of Quantity as Empirical Hypotheses

Relying on the classical view, the question of *why* we can measure some attributes can be answered in a straightforward manner: some attributes are measurable because they possess quantitative structure. The next question we must ask is: *how* can we decide whether or not an attribute is quantitative? The method of finding out if some attribute satisfies conditions of quantity is part of what Michell (1997, 1999) calls the scientific task. (For the sake of completeness it should be mentioned that this is only the first part of the task of quantification. The second part is the instrumental task, which consists of the development of measurement instruments. Obviously the scientific task must always precede the instrumental task.) It must be stressed that 'the hypothesis that some attribute is quantitative is a quite specific hypothesis, one never logically necessary' (Michell, 1999, p. 67). That is, *quantitative structure can be ascribed to an attribute only if it empirically satisfies the conditions of quantity*. The scientific task therefore always implies testing empirical hypotheses.

The first and therefore most basic condition of quantity structure demands that '*any two magnitudes of the same quantity are either identical or different*' (Michell, 1999, p. 52). Let me take for the purpose of illustration conjoint measurement theory, which applies to situations where it is assumed that an attribute is affected by two or more components. 'A simple example is the attribute momentum which is exhibited by physical objects and which is affected both by their mass and by their velocity' (Krantz et al., 1971, p. 245). It is often ignored that in order to make use of this theory the objects investigated must be classifiable according to the values (or levels of magnitude) the components possess (Krantz et al., 1971, p. 246; Michell, 1990, p. 69; 1999, p. 206). That is, we must be able to identify objects of equal mass and equal velocity.

It will be sufficient to focus solely on the first condition of quantity in order to decide if psychological attributes are measurable, since it is a *necessary* (though of course not sufficient) requirement for the identification of quantitative structure. Moreover, notice that conditions of quantity stand in relations of *hierarchical dependence*, which implies that we can very well empirically satisfy 'lower' conditions of quantity (e.g., relations of equivalence) without taking into account conditions of 'higher' order (i.e., relations of order or additivity). Conversely, if we cannot satisfy subordinate conditions, the satisfaction of the superordinate conditions is logically excluded.

We must also take into account that relations between magnitudes do not always manifest themselves directly in relations between objects which are the assumed bearers of the magnitudes, because, as Michell (1999) explains, 'the observable equivalence relation between objects (e.g., the fact that two marbles perfectly balance one another) and the identity relation between magnitudes (e.g., the fact that the weight of two marbles is the same) are logically distinct' (p. 70). In view of this, it is useful to differentiate between extensive and intensive quantities. Extensive quantities (e.g., length) are those for which the additivity of the quantity 'is evident to us more or less directly from the behavior of some objects manifesting magnitudes of the quantity' (Michell, 1999, p. 54). With regard to intensive quantities (e.g., temperature, pressure, electrical tension, etc.), relations of additivity are not directly observable and therefore in principle not amenable to direct testing. Actually, in this case not even the equivalence relation between magnitudes is directly testable. In short, *if it comes to intensive quantities, the conditions of quantity can be tested only indirectly*, that is, by means of an associated observable attribute in the manner exemplified in the next section. In psychology such relations are commonly conceptualized as relations between observables (or manifest variables) and theoretical or hypothetical constructs (or latent variables). Since psychological attributes can be thought of only as intensive quantities, we can restrict the question of their measurability to asking how we can find out empirically if intensive quantities satisfy the first condition of quantity.

Testing the First Condition of Quantity

How can we test the first condition of quantity? More precisely, how are we to proceed if we want to identify equal levels of intensive quantities? It might be useful to resort to an example where this task has already been accomplished successfully. Let us take for instance Ohm's law (1826), which specifies the quantitative relation between tension, V, electrical resistance, R, and intensity of the electrical current, I. In order to get started, Ohm had to combine the *same* level of V with *different* levels of R (and the other way round) and observe the effect of this combination on I. For brevity I will describe only how he dealt with V. He solved the problem by making use of the empirical fact discovered by Oersted that a freely turning magnetic needle attached to a string is deflected by the presence of an electric current in a nearby conductor (Bordeau, 1982). That is, Ohm used the degree of deflection, *D*, as an indicator for the degree of strength of *V*. The observable is in this case already a measurable extensive quantity, and equal levels of *D* can therefore easily be determined. The logic behind this strategy is: whenever the needle rests at a certain angle and if no other factors (such as the material composition of the resistors, the magnetic field of the earth, the tension in the string to which the magnetic needle is attached, etc.) systematically disturb the investigated relation, we can confidently conclude that *V* is constant at a determinate level of magnitude. This, in principle, is the method for determining equal levels of intensive quantities.

Notice, however, that we cannot simply make the inference that if the observable attribute is equal in level of magnitude, then the corresponding hypothetical construct must also be equal in level of magnitude. For instance, it is often assumed in psychology that persons with an equal total score on some test are equal in ability. Such inference is not permissible without further empirical justification, since, as Ohm's (1826) case illustrates, equal levels of *D* might very well be determined by different levels of *V* in combination with some other factors also influencing *D*. That is, *we cannot take for granted that equal levels of some manifest variable necessarily correspond to equal levels of some latent variable, but we must ascertain by experiment that this really is the case*.

In conclusion, the procedure to quantify intensive quantities consists, firstly, in varying the hypothetical construct in intensity or keeping it constant at some level and, secondly, in observing the effect the manipulation has on an observable. We must also make sure that no disturbances interfere with the relations investigated. (I will return to this important point in some detail in the next section.) Actually this method corresponds to what Mill (1843/1974) refers to as 'the method of concomitant variations' (pp. 398–403). It is considered the most efficient and most often used method of causal analysis (Cook & Campbell, 1979; Tetens, 1987). Furthermore, although we cannot observe levels of the theoretical construct directly, we must nevertheless be able to manipulate it in intensity *independently* of the observable; otherwise we will not be able to apply the method described (cf. Johnson, 1936, 1945, 1954). Ohm, for instance, manipulated *V* with the help of a thermocouple, which is a device that relies on the empirical fact that the junctions of dissimilar metals, when maintained at different temperatures, produce a constant flow of current through an electrical circuit (Bordeau, 1982).

Experiment, Apparatus and Disturbances

We have just seen *how* physicists proceed in order to quantify intensive quantities. Let us now focus on *what* really makes them so successful in solving the scientific task. First of all note that Ohm (1826) solved it by means of experiment. This is no exception but rather the rule. As a matter of fact in physics the task of quantification usually involves the use of experiment. In his *General Considerations Concerning Scientific Apparatus*, Maxwell (1890/1965) explains:

> In designing an Experiment the agents and phenomena to be studied are marked off from all others and regarded as the Field of Investigation. All agents and phenomena not included within this field are called Disturbing Agents, and their effects Disturbances; and the experiment must be so arranged that the effects of these disturbing agents on the phenomena to be investigated shall be as small as possible. (p. 505)

In short, there are two important uses of experiment. Firstly, in experiment we *manipulate* agents (or causes) and observe how this manipulation affects phenomena (or effects). Secondly, in experiment we *control* disturbances. For example:

> … in experiments where we endeavor to detect or to measure a force by observing the motion which it produces in a movable body, we regard Friction as a disturbing agent, and we arrange the experiment so that the motion to be observed may be impeded as little as possible by friction. (Maxwell, 1890/1965, p. 506)

Before I go further into details about the role of experiment for solving the scientific task, let me add here some important remarks about the nature of disturbances. Firstly, disturbances originate in the causal complexity of the world. Secondly, there are basically two kinds of disturbances: random and systematic disturbances. Methods suited for dealing with random disturbances are not suited for treating systematic disturbances, and vice versa. That is, while random error can be dealt with by means of some statistical theory (e.g., the Gaussian theory of errors), there is no general theoretical solution for the problems posed by systematic error (Taylor, 1997). The only way to deal with the influence of systematic disturbances is by controlling it in experiment. For example, we can try to keep it constant or isolate it (Parthey & Wahl, 1966, pp. 167–174). Systematic disturbances might also be included in the field of investigation, in which case they are phenomena to be studied in their own right. Note that there is no essential difference between phenomena and disturbances. In nature there are only causes and effects. Of course, the influence of systematic disturbances usually cannot be entirely controlled. However, it is sufficient to reduce it to the level of random disturbances, at which point they can be dealt with by means of error theory. In conclusion, *the problems posed by systematic disturbances are not theoretical problems. They can only be solved by experimental—not theoretical—means.*

Consequently, if it comes ‘to questions of how to apply the abstract measurement principles to fallible or incomplete data’ (Cliff, 1992, p. 189), we must always take into account both kinds of error. In particular, with regard to the application of measurement theory in psychology, problems posed by disturbances cannot be dealt with indiscriminately by theoretical means; nor can they be dealt with by probabilistic theories of measurement (Bradley & Terry, 1952; Rasch, 1960/1980; Thurstone, 1927, 1959), by random additive conjoint measurement (Falmagne, 1976, 1978), by algebraic or probabilistic theories of thresholds (Block & Marschack, 1960; Fechner, 1860; Fishburn, 1973; Lipps, 1906; Luce, 1956), or by any similar theoretical approach (Karabatsos, 2001, 2005, 2006; Scheiblechner, 1995, 1999). In view of systematic errors, all measurement theories, whether construed as deterministic or probabilistic, are on equal footing and the problems described subsequently apply to all without exception.

Furthermore, it would be imprudent to treat the problem of systematic disturbances lightly. Consider that in order to apply measurement theory one needs not only a way of identifying values of the attributes involved, but also 'some way of, first, identifying and, then, controlling other relevant causes [i.e., systematic disturbances], so that the features of the data diagnostic of additive [i.e., quantitative] structure are not swamped by error' (Michell, 1999, pp. 206–207). In short, *the failure to control systematic disturbances renders the discovery of quantitative structure impossible*. At some point Ohm, for example, observed that the strength of the electrical current decreases rapidly over time. He rightly suspected that the cause for the loss of strength was due to the voltaic cells used as source of electricity. Only after replacing them with a thermocouple was he able to control the systematic disturbance caused by the source of electricity and thereby attain the required constant tension in the electrical circuit (Ohm, 1826, pp. 139–144; see also Teichmann, 1977). If, on the other hand, he would not have been able to control this or other systematic disturbances he would not have been able to establish his law. Indeed, the art of experimentation consists to a considerable extent of the identification and control of disturbing factors.

Let me now return to the role of the experiment for the discovery of quantitative laws. It is often not appropriately understood that in the presence of systematic disturbances the particular kind of experiment described by Maxwell is unavoidable (for an overview of types of experiment, see Parthey & Wahl, 1966, chap. 7). That is, what is required in this case is the construction of experimental apparatuses or machines. Notice also that the use of apparatus presupposes a deliberate act of design and construction. In general we don't use natural objects (i.e., objects as we find them in nature) for the purpose of apparatus construction; we usually have to tool and machine them first. The task of apparatus construction therefore typically involves the use of artificial objects (Dingler, 1921; Tetens, 1987). To give another example: it is seldom acknowledged that the second law of motion, which is a statement relating force, mass and acceleration, was confirmed empirically some time after Newton (1687/1999) formulated it. It was George Atwood (1784) who made up a simple but nevertheless ingenious apparatus to verify it (Hanson, 1965, pp. 99–105). Strictly speaking, until Atwood's experiment, Newton's second law was just a quantitative hypothesis. The modern version of the apparatus, now known as Atwood's machine, connects two objects of different mass by a string passing over a pulley. None of the components of the Atwood machine are natural objects. The pulley is a wheel with a groove along its edge, for holding the string. It must be worked in such a way that it is as round as possible. Moreover, the groove must be machined so that the friction it causes is reduced to a minimum. The string must be manufactured of such a material that the tension is the same all along its length, which implies that the string is as inelastic as possible. The objects accelerated are all manufactured of the same material, are of regular geometrical shape, and so on. Certainly, apparatus must be installed in a laboratory in order to shield the experiment against atmospheric factors like temperature, humidity, wind, and so forth.

Finally, as explained above, if the data are superimposed by systematic error, no quantitative structure can be discovered. Under such adverse circumstances the scientific task of quantification can be solved only through the construction of apparatus. This is actually a matter of definition. For, under causally intricate circumstances quantification can be obtained only by intervention on the part of the experimenter, in the manner exemplified by Ohm's (1826) apparatus or Atwood's (1784) machine. The action of disjoining an undifferentiated and interwoven bundle of causal relations, as usually offered by nature, and systematically joining together the cause–effect relations of interest, while controlling disturbing factors, is what the construction of an apparatus essentially consists of. That is, the result of the activity of 'taking apart' and 'putting together' is what is *called* 'experimental apparatus'. Hence, *against the causal complexity of the world the measurability of quantitative attributes cannot be discovered without the help of apparatus*. (Of course, not all experiments involve the use of apparatus and not all apparatuses are constructed for the purpose of the discovery of quantitative laws or even for an experimental or scientific purpose.) There is just one exception to the rule: if nature spontaneously behaves like an apparatus 'we may discover, by mere observation without experiment, a real uniformity in nature' (Mill, 1843/1974, p. 386). Or as Cartwright (1999) states: 'Sometimes the arrangement of the components and the setting are appropriate for a law to occur naturally, as in the planetary system; more often they are engineered by us, as in a laboratory experiment' (p. 49; see also Boumans, 2005).

The Galilean Revolution

Relying on the previous sections, I will explain now what I mean by the Galilean revolution. Galileo was the first scientist to realize that in order to discover laws of nature it might not be sufficient to rely on the method of *passive* observation, as supporters of Aristotelian physics believed. Rather, one must *actively* intervene in the course of nature and deliberately manipulate the phenomena of interest. Galileo also clearly recognized that the divergence between some theoretical idealization, like his law of free fall, and ‘the intractable irregularity of the real world’ (McMullin, 1985, p. 248) doesn’t necessarily invalidate the former. That is, contrary to his Aristotelian opponents, he did not accept that because of divergences between theory and empirical observation, we must give up the idea that ‘the book of nature’ is written in the language of mathematics, because, as he emphasized, they may be due to impediments and irregularities active in experimentally unconstrained nature (for details see McMullin, 1985). Galileo also was the first to recognize that we must consider two kinds of disturbances: some which are so ‘large’ that they interfere to such an extent that they impede the discovery of laws of nature and others which are so ‘small’ that they don’t endanger the process of discovery (Koertge, 1977). But most importantly, he realized that disturbances can be reduced or removed through the construction of apparatus; that through apparatus construction, the different causal threads active in nature can be taken apart and investigated separately. As a matter of fact, he constructed the first such apparatus, an inclined plane, to empirically verify his quantitative hypothesis about the law of free fall (Galilei, 1638/1991, pp. 178–179). The change in scientific practice initiated by Galileo marked a revolutionary turn in the natural sciences. The Aristotelian view which had until then dominated academic physics was now surpassed by this new philosophy of experimentation. Aristotelian physicists rejected the idea that laws of nature can be discovered with the help of machines. They argued that what are revealed through apparatus construction are not really empirical laws, since the observed regularities are forced upon nature by mechanical means whereby nature is, so to speak, outwitted instead of being really deciphered (Mittelstrass, 1970, pp. 172–174; 1972).

Despite its importance, the role of experiment, and of experimental apparatus in particular, has until recently gone almost unnoticed in the philosophy of science. The earliest extensive philosophical appreciation of the use of experiment is to my knowledge Hugo Dingler’s *Das Experiment* (1928). Dingler’s thoughts were received and refined by methodological constructivism (Butts & Brown, 1989; Janich, 1985, 1997; Lorenzen, 1987, Tetens, 1987). The importance of apparatus has been rediscovered independently of this development in the context of so-called ‘new experimentalism’ (Ackermann, 1989; Franklin, 1986, 1990; Galison, 1987; Gooding, Pinch, & Schaffer, 1989; Heidelberger & Steinle, 1998; Radder, 2003).

Reflection on the role of apparatus led some philosophers of science to question central aspects of the traditional, realist view about laws of nature. For instance, Cartwright (1999) challenges the basic assumption of realism that laws of nature are universally true (i.e., apply everywhere in the universe). She argues that almost all physical laws are *ceteris paribus* laws since they are obtained only under controlled laboratory conditions and are therefore true only under these experimentally constrained circumstances. Hacking (1983) questions another aspect of the realist view, namely the idea that laws of nature exist independently of experiment. He claims that within experiment, phenomena are not discovered but created. According to him, this doesn't imply that scientists can create phenomena at will, as someone might object, but that they did not exist out there in nature before their realization in experiment. A similar, anti-realist view was already supported by Dingler much earlier; most emphatically in his lecture *Der Glaube an die Weltmaschine und seine Überwindung* (1932). He also argued that laws of nature are not really a matter of discovery but of making. Some of his disciples, like Janich (1978) and Tetens (1987), also vigorously denounce the traditional view as misrepresenting the facts. Instead of laws of nature they prefer to speak of *Apparategesetze* ('apparatus laws'), since, strictly speaking, the regularities created in experiment apply only to and are only valid of experimental apparatus. But as Kroes (2003) recently demonstrated, as a rejoinder to Hacking's challenge, realism can account just as well for what is happening in an experiment (cf. also Franklin, 1984).

In my view both positions, although apparently contradictory, are logically valid descriptions of experimental practice; but, unfortunately, given limited space I cannot go further into details. For my purpose it is only important to realize that none of the aforementioned anti-realist philosophers dispute the necessity and usefulness of the experimental method. Most importantly, they don't question that 'laws of nature' are usually established under laboratory conditions with the help of experimental apparatus. Accordingly, they don't propose to change or even improve what scientists are really doing, but want to offer what they believe is a more accurate account of the epistemology of apparatus. For these reasons the conclusions with regard to the possibility of measurement in psychology are not affected by which philosophy of apparatus one prefers. That I make use of the traditional view as a theoretical framework for my critique is solely a matter of convenience and as such is determined by the fact that realism is the dominant view in the natural sciences in general and in psychology in particular. I could just as well have formulated my objections in the anti-realist language of constructivism (Trendler, 2006).

The Millean Quantity Objection

As explained above, in the face of causally intricate circumstances the scientific task of quantification requires the construction of apparatus. Only through the construction of experimental apparatus is it possible under these circumstances to manipulate the attributes of interest and control disturbances to such an extent as to allow testing conditions of quantity. Conversely, if the relevant factors cannot be handled through apparatus construction, the task of quantification is not solvable. With regard to psychology this implies:

if psychological phenomena are not dependent or cannot be made to depend on a manageable set of conditions, then they are not measurable.

Let me call this the Millean quantity objection, since, to my knowledge, it was John Stuart Mill (1843/1974, pp. 379–387) who initially advanced it. Although he does not explicitly argue that the quantification of psychological attributes is impossible, the objection can easily be deduced from his argumentation. First, he states that in order to discover laws of nature we may have recourse 'either to observation or to experiment; we may either *find* an instance in nature suited to our purposes, or, by an artificial arrangement of circumstances, *make* one' (p. 381). Following this, he points out that if it is not possible to apply either of the two methods, no laws of nature may be discovered. As an exemplar of such unfavorable situations he mentions psychological phenomena. He draws attention to the fact

> … that in every instance in which we see a human mind developing itself, or acting upon other things, we see it surrounded and obscured by an indefinite multitude of unascertainable circumstances, rendering the use of the common experimental methods almost delusive. (p. 384)

Considerable time has passed since Mill formulated his objection, so one might wonder if it still applies today. Maybe revolutionary developments in measurement theory render it obsolete? On the contrary, as I will argue in the next section, the substantial progress made in measurement theory mainly in the second half of the last century makes the objection even more compelling.

The Case against Measurement in Psychology

To begin with, the method of observation is, as previously argued, suited for the discovery of quantitative structure only if circumstances in nature are such that the influence of systematic disturbances is negligible. This is obviously not the case in psychology. There is also no dispute among psychologists that, just as in physics, systematic disturbances must be accounted for by means of experimental control (Goodwin, 1995; Nunnally, 1978; Pedhazur & Pedhazur-Schmelkin, 1991; Rosenthal & Rosnow, 1991). Hence, in order to test the hypothesis that psychological attributes are quantitative, there is no alternative but to make use of experiment.

There can also be no doubt that psychological phenomena can be manipulated and controlled to some extent. However, it is far from evident that they are manageable to the degree required by measurement theory. For example, most psychologists will agree that motivation can be manipulated by means of different amount of reward (e.g., amount of money). But, as explained above, in order to begin solving the scientific task we would have to test the first condition of quantity: that is, we would have to determine motivation in test subjects in such a way that we obtain *equal levels* of motivation in the same subject (or in different subjects) for over the time of the experiment. As also already explained, we cannot take for granted that equal levels of an observable necessarily correspond to equal levels of the associated hypothetical construct. That is, equal amounts of reward might not automatically lead to equal levels of motivation. In order to identify equal levels of motivation we would have to make use, for instance, of the method of concomitant variation, as exemplified by means of Ohm's (1826) experiment. If we assume in addition that there is a causal relation between motivation and the reaction time for some test items, then the result we must ultimately aim at is: if the same amount of reward is applied and if no systematic disturbances interfere, then the resulting reaction time must be equal in value, in the limits of random errors, over experimental replications. Only if this criterion is empirically satisfied can we confidently conclude that equal amounts of reward generate equal levels of motivation.

But isn't this a utopian objective? Admittedly, if testing the conditions of quantity fails, it does not automatically follow that there is no quantitative structure to be discovered; we might reasonably conjecture that systematic disturbances have impeded the discovery. However, if one opts for the second alternative, one would have to seriously contemplate how the control of the assumed disturbances can be accomplished. For instance, one would have to make sure that no other factors (e.g., ability, learning, attention, etc.) influence performance in a systematic way. Control could be achieved by keeping disturbances constant at a certain level over the time of the experiment. Unfortunately this strategy confronts us with the same problem all over again, since how should we identify, say, constant levels of ability? Another option would be to isolate motivation from the influence of ability and of the other systematic disturbances. Obviously this is even less practicable.

But, it might be argued that although the resources of psychology are indeed insufficient to solve the scientific task, progress in other substantive areas like neurophysiology may provide the necessary access to psychological phenomena. If we adopt this approach we must of course take the neurophysiological structure of the brain as the material substrate of psychological phenomena (cf. Krech, 1950). Furthermore, notice that this approach is only promising if psychological attributes are at least clearly localizable; otherwise we would be confronted with the problem of causal entanglements again. Unfortunately, the hypothesis of distinct cortical localization (Finger, 1994, chap. 3) has been clearly refuted, since, as modern neurosciences inform us, the brain is a dynamically adaptive and flexible system (Cabeza & Kingstone, 2006; Crick, 1979; Fuillet, Dufour, & Pelletier, 2007; Kandel, Schwartz, & Jessell, 2006; Kertesz, 1994; Mesulam, 2000; Nicolelis & Ribeiro, 2006; Posner, Petersen, Fox, & Raichle, 1988; Raichle, 1993). Hence, at present it is certainly not possible to identify specific levels of magnitudes of psychological attributes simply by observing the working of the brain. Furthermore, in my view it is also unlikely that present knowledge about the dynamism and holism of brain activity will be reversed in the future in favor of some strict localization or that the brain will turn out to be such simple 'apparatus' as required by measurement theory.

Actually, and most importantly, in face of the causal complexity of brain activity we would have to consider applying the method adopted in physics to deal with such problems, namely the Galilean method of apparatus construction. That is, we would have to envisage an active intervention in the sense of 'straightening the course of nature' by disjoining the relevant causal threads and rejoining only those of interest, so that optimal conditions are set up for quantitative relations to manifest themselves. But, to return to the example above, and putting aside ethical concerns and the problem of identifying neurological representations of psychological phenomena: how should we 'slice and dice' the brain of a test subject in such a way that only motivation influences reaction time and that all other factors which might additionally influence behavior are under control? Obviously the Galilean procedure is inapplicable in psychology.

In conclusion: *psychological phenomena are not sufficiently manageable. That is, they are neither manipulable nor are they controllable to the extent necessary for an empirically meaningful application of measurement theory. Hence they are not measurable*. In my view no substantial progress will be reached in psychology until we accept psychological phenomena as they really are, namely in their natural 'muddled' state. It might be cold comfort, but physicists would find themselves in the same hopeless situation if they were not to be allowed to construct apparatus (see Cartwright, 1999, chap. 3).

Conclusions

In essence, I have argued that in psychology the extremely successful Galilean method reaches the limits for its successful application. The problem is not that psychological systems are more complex than physical systems; they might be, but the crucial difference is that, contrary to physical phenomena, psychological phenomena cannot be *made* to depend on a small set of manageable conditions. In other words, the very effective method used in physics of manipulating and controlling phenomena through apparatus construction is not applicable in psychology. This difference explains in my view the success of quantification in physics since Galileo and conversely the failure of similar attempts in psychology since Fechner, and this is also the reason why I believe that the Galilean revolution never happened in psychology.

In his commentary of the publication of the last two volumes of *Foundations of Measurement* (Luce, Krantz, Suppes, & Tversky, 1990; Suppes, Krantz, Luce, & Tversky, 1989), Norman Cliff (1992) points out that representational measurement theory, and conjoint measurement theory in particular, had no impact on psychology. He therefore calls conjoint measurement 'the revolution that never happened' (p. 186). For different reasons he also is not optimistic that the situation will ever change. But Michell (1999) resists this pessimistic outlook and insists in calling conjoint measurement 'the revolution that happened' (p. 193). In my opinion this dispute is beside the point. Conjoint measurement theory might be a revolution that happened in measurement theory, but it is actually irrelevant to psychology. It is the Galilean revolution psychologists would have to unfold. Unfortunately, as it turns out, it is the revolution that *cannot* happen.

We must now also recognize that Norman Campbell (1920, 1928) was right after all about the non-measurability of psychological attributes. But he was right for the wrong reason. Campbell and those following him (e.g., Guild, 1938) mistakenly believed that the only form of fundamental measurement is by means of the empirical concatenation operation, a conviction Campbell (1928) expressed in his second law of measurement. But, as conjoint measurement theory demonstrates, even if the operation of concatenation is not applicable to psychological attributes, it doesn't necessarily follow that they are fundamentally non-measurable (Michell, 1990). From my point of view the optimism which might be accorded to this statement is highly exaggerated. In psychology we cannot even satisfy Campbell's first law of measurement (see Campbell, 1928, chap. 1), which contains the demand for equivalence between magnitudes. Paradoxically Campbell was not too restrictive with regard to psychology, but too lenient. It is remarkable that attempts at quantification in psychology fail at such an early stage.

The reader might wonder what the implications are for psychology as a quantitative science. It has repeatedly been stressed that the use of classical test theory and of the so popular linear models of statistical analysis (e.g., analysis of variance, linear regression or factor analysis) requires that data are quantitative (Fischer, 1968, 1974; Sixtl, 1982, 1985, 1998; Wright, 1997; Wright & Linacre, 1989; Wright & Masters, 1982). This cannot be overstated, since, as Wright (1997) notes, ignoring the quantitative imperative 'is why so much social science has turned out to be no more than transient description of never-to-be-reencountered situations, easy to contradict with almost any replication' (p. 35; see also Barrett, 2008). Accordingly, measurement theory is thought to offer a remedy against this regrettable state of affairs. Unfortunately, the conclusion reached here is that it is not. That is, the application of measurement theory, irrespective whether it is construed as deterministic or probabilistic, is also not relevant to achieving substantial progress in psychology. Other, more suited methods for the domain of psychology must be found. It might therefore be wise to seriously reconsider Johnson's recommendation: 'Those data should be measured which can be measured; those which cannot be measured should be treated otherwise. Much remains to be discovered in scientific methodology about valid treatment and adequate and economic description of non-measurable facts' (Johnson, 1936, p. 351).

References

Ackermann, R. (1989). The new experimentalism. *British Journal for the Philosophy of Science*, *40*, 185–190.

Andrich, D. (1988). *Rasch models for measurement*. Newbury Park, CA: Sage.

Andrich, D. (2004). Controversy and the Rasch model. *Medical Care*, *42*(1 Suppl. 1), 7–16.

Atwood, G. (1784). *A treatise on the rectilinear motion and rotation of bodies; with a description of original experiments relative to the subject*. Cambridge, UK: Archdeacon.

Barrett, P. (2008). The consequence of sustaining a pathology: Scientific stagnation. A commentary on the target article 'Is psychometrics a pathological science?' by Joel Michell. *Measurement*, *6*, 78–123.

Berka, K. (1983). *Measurement: Its concepts, theories and problems*. Dordrecht, The Netherlands: Reidel.

Blais, M.A. (2003). Have you heard we're having a revolution? The coming of modern test theory. *Journal of Personality Assessment*, *80*, 208–210.

Blinkhorn, S.F. (1997). Past imperfect, future conditional: Fifty years of test theory. *British Journal of Mathematical and Statistical Psychology*, *50*, 175–185.

Block, H.D., & Marschak, J. (1960). Random orderings and stochastic theories of responses. In I. Olkin, S. Ghurye, H. Hoeffding, W. Madow, & H. Mann (Eds.), *Contributions to probability and statistics* (pp. 97–132). Stanford, CA: Stanford University Press.

Bordeau, S.P. (1982). *Volts to Hertz: The rise of electricity*. Minneapolis, MN: Burgess.

Boumans, M. (2005). Measurement outside the laboratory. *Philosophy of Science*, *72*, 850–863.

Bradley, R.A., & Terry, M.E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, *39*, 324–345.

Brown, J.F. (1934). A methodological consideration of the problem of psychometrics. *Erkenntnis*, *4*, 46–61.

Butts, R.E., & Brown, J.R. (Eds.). (1989). *Constructivism and science*. Dordrecht, The Netherlands: Kluwer.

Cabeza, R., & Kingstone, A. (Eds.). (2006). *Handbook of functional neuroimaging of cognition* (2nd ed.). Cambridge, MA: MIT Press.

Campbell, N. (1920). *Physics: The elements*. Cambridge, UK: Cambridge University Press.

Campbell, N. (1928). *An account of the principles of measurement and calculation*. London: Longmans.

Cartwright, N. (1999). *The dappled world*. Cambridge, UK: Cambridge University Press.

Cliff, N. (1992). Abstract measurement theory and the revolution that never happened. *Psychological Science*, *3*, 186–190.

Cook, T.D., & Campbell, D.T. (1979). *Quasi-experimentation*. Boston: Houghton Mifflin.

Crick, F.H.C. (1979). Thinking about the brain. *Scientific American*, *241*(3), 219–232.

Dingler, H. (1921). *Physik und Hypothese* [Physics and hypothesis]. Berlin, Germany: Walter de Gruyter.

Dingler, H. (1928). *Das Experiment* [The experiment]. Munich, Germany: Reinhardt.

Dingler, H. (1932). *Der Glaube an die Weltmaschine und seine Überwindung* [Overcoming the belief in the world-machine]. Stuttgart, Germany: Enke.

Ellis, B. (1966). *Basic concepts of measurement*. Cambridge, UK: Cambridge University Press.

Falmagne, J.-C. (1976). Random conjoint measurement and loudness summation. *Psychological Review*, *83*, 65–79.

Falmagne, J.-C. (1978). A representation theorem for finite random scale systems. *Journal of Mathematical Psychology*, *18*, 52–72.

Fechner, G.T. (1860). *Elemente der Psychophysik* [Elements of psychophysics] (Vol. 1). Leipzig, Germany: Breitkopf & Härtel.

Finger, S. (1994). *Origins of neuroscience*. New York: Oxford University Press.

Fischer, G.H. (1968). Neue Entwicklungen in der psychologischen Testtheorie [New developments in psychological test theory]. In G.H. Fischer (Ed.), *Psychologische Testtheorie* [Psychological test theory] (pp. 15–158). Bern, Switzerland: Huber.

Fischer, G.H. (1974). *Einführung in die Theorie psychologischer Tests* [Introduction to the theory of psychological tests]. Bern, Switzerland: Huber.

Fishburn, P.C. (1973). Binary choice probabilities: On the varieties of stochastic transitivity. *Journal of Mathematical Psychology*, *10*, 327–352.

Franklin, A. (1984). The epistemology of experiment. *British Journal for the Philosophy of Science*, *35*, 381–390.

Franklin, A. (1986). *The neglect of experiment*. Cambridge, UK: Cambridge University Press.

Franklin, A. (1990). *Experiment, right or wrong*. Cambridge, UK: Cambridge University Press.

Fuillet, L., Dufour, H., & Pelletier, G. (2007). Brain of a white-collar worker. *The Lancet*, *370*(9583), 262.

Galilei, G. (1991). *Dialogues concerning two new sciences*. Amherst, NY: Prometheus. (Original work published 1638)

Galison, P. (1987). *How experiments end*. Chicago: University of Chicago Press.

Gooding, D., Pinch, T., & Schaffer, S. (Eds.). (1989). *The uses of experiment*. Cambridge, UK: Cambridge University Press.

Goodwin, C.J. (1995). *Research in psychology: Methods and design*. New York: Wiley.

Guild, J. (1938). Are sensation intensities measurable? *British Association for the Advancement of Science*, *108*, 296–328.

Hacking, I. (1983). *Representing and intervening*. Cambridge, UK: Cambridge University Press.

Hanson, N.R. (1965). *Patterns of discovery: An inquiry into the conceptual foundations of science*. Cambridge, UK: Cambridge University Press.

Heidelberger, M., & Steinle, F. (Eds.). (1998). *Experimental essays*. Baden-Baden, Germany: Nomos.

Helmholtz, H. von. (1998). Zählen und Messen, erkenntnistheoretisch betrachtet [Numbering and measuring from an epistemological viewpoint]. In E. Bonk (Ed.), *Schriften zur Erkenntnistheorie* [Epistemological writings] (pp. 99–129). Vienna: Springer. (Original work published 1887)

Hölder, O. (1901). Die Axiome der Quantität und die Lehre vom Mass [The axioms of quantity and the theory of measurement]. *Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig, Mathematisch-Physische Classe*, *53*, 1–64.

Hölder, O. (1924). *Die mathematische Methode* [The mathematical method]. Berlin, Germany: Springer.

Janich, P. (1978). Physics, natural science or technology. In W. Krohn, E.T. Layton, & P. Weingart (Eds.), *The dynamics of science and technology* (pp. 3–27). Dordrecht, The Netherlands: Reidel.

Janich, P. (1985). *Protophysics of time*. Dordrecht, The Netherlands: Reidel.

Janich, P. (1997). *Das Maß der Dinge* [The measure of things]. Frankfurt am Main, Germany: Suhrkamp.

Johnson, H.M. (1936). Pseudo-mathematics in the mental and social sciences. *American Journal of Psychology*, *48*, 342–351.

Johnson, H.M. (1945). Are psychophysical problems genuine or spurious? *American Journal of Psychology*, *58*, 189–211.

Johnson, H.M. (1954). On verifying hypotheses by verifying their implicates. *American Journal of Psychology*, *67*, 723–727.

Kandel, E.R., Schwartz, J.H., & Jessell, T.M. (Eds.). (2006). *Principles of neural science* (5th ed.). New York: McGraw-Hill.

Karabatsos, G. (2001). The Rasch model, additive conjoint measurement, and new models of probabilistic measurement theory. *Journal of Applied Measurement*, *2*, 389–423.

Karabatsos, G. (2005). The exchangeable multinomial model as an approach to testing deterministic axioms of choice and measurement. *Journal of Mathematical Psychology*, *49*, 51–69.

Karabatsos, G. (2006). Bayesian nonparametric model selection and model testing. *Journal of Mathematical Psychology*, *50*, 123–148.

Kertesz, A. (Ed.). (1994). *Localization and neuroimaging in neuropsychology*. San Diego, CA: Academic Press.

Koertge, N. (1977). Galileo and the problem of accidents. *Journal of the History of Ideas*, *38*, 389–408.

Krantz, D.H., Luce, R.D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement: Vol. 1. Additive and polynomial representations*. New York: Academic Press.

Krech, D. (1950). Dynamic systems, psychological fields, and hypothetical constructs. *Psychological Review*, *57*, 283–290.

Kroes, P. (2003). Physics, experiments, and the concept of nature. In H. Radder (Ed.), *The philosophy of scientific experimentation* (pp. 68–86). Pittsburgh, PA: University of Pittsburgh Press.

Lewin, K. (1931). The conflict between Aristotelian and Galileian modes of thought in contemporary psychology. *Journal of Genetic Psychology*, *5*, 141–177.

Lipps, G.F. (1906). *Die psychischen Massmethoden* [Psychological measurement methods]. Brunswick, Germany: Vieweg.

Lorenzen, P. (1987). *Constructive philosophy*. Amherst: University of Massachusetts Press.

Luce, R.D. (1956). Semiorders and a theory of utility discrimination. *Econometrica*, *24*, 178–191.

Luce, R.D., Krantz, D.H., Suppes, P., & Tversky, A. (1990). *Foundations of measurement: Vol. 3. Representations, axiomatization, and invariance*. New York: Academic Press.

Luce, R.D., & Suppes, P. (2002). Representational measurement theory. In H. Pashler & J. Wixted (Eds.), *Stevens' handbook of experimental psychology* (3rd ed., Vol. 4, pp. 1–41). New York: Wiley.

Luce, R.D., & Tukey, J.W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, *1*, 1–27.

Maxwell, J.C. (1965). General considerations concerning scientific apparatus. In W.D. Niven (Ed.), *The scientific papers of James Clerk Maxwell* (Vol. 2, pp. 505–522). New York: Dover. (Original work published 1890)

McMullin, E. (1985). Galilean idealization. *Studies in History and Philosophy of Science*, *16*, 247–273.

Mesulam, M.-M. (Ed.). (2000). *Principles of behavioral and cognitive neurology* (2nd ed.). Oxford, UK: Oxford University Press.

Michell, J. (1990). *An introduction to the logic of psychological measurement*. Hillsdale, NJ: Erlbaum.

Michell, J. (1997). Quantitative science and the definition of measurement in psychology. *British Journal of Psychology*, *88*, 355–383.

Michell, J. (1999). *Measurement in psychology*. Cambridge, UK: Cambridge University Press.

Michell, J. (2000). Normal science, pathological science and psychometrics. *Theory & Psychology*, *10*, 639–667.

Michell, J. (2004). Item response models, pathological science and the shape of error: Reply to Borsboom and Mellenbergh. *Theory & Psychology*, *14*, 121–129.

Mill, J.S. (1974). *A system of logic: Ratiocinative and inductive*. Toronto, ON: University of Toronto Press. (Original work published 1843)

Mittelstrass, J. (1970). *Neuzeit und Aufklärung* [Modern times and enlightenment]. Berlin, Germany: Walter de Gruyter.

Mittelstrass, J. (1972). The Galilean revolution. *Studies in the History and Philosophy of Science*, *2*, 297–328.

Narens, L. (1985). *Abstract measurement theory*. Cambridge, MA: MIT Press.

Narens, L., & Luce, R.D. (1986). Measurement: The theory of numerical assignments. *Psychological Bulletin*, *99*, 166–180.

Newton, I. (1999). *Mathematical principles of natural philosophy*. Berkeley: University of California Press. (Original work published 1687)

Nicolelis, M.A.L., & Ribeiro, S. (2006). Seeking the neural code. *Scientific American*, *295*(6), 70–77.

Nunnally, J.C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Ohm, G.S. (1826). Bestimmung des Gesetzes, nach welchem Metalle die Contaktelektricität leiten, nebst einem Entwurfe zu einer Theorie des voltaischen Apparats und des Schweiggerschen Multiplicators [Determination of the law in accordance with which metals conduct contact electricity, together with an outline of a theory of the voltaic apparatus and of Schweigger's multiplier]. *Journal für Chemie und Physik*, *46*, 137–166.

Parthey, H., & Wahl, D. (1966). *Die experimentelle Methode in Natur- und Gesellschaftswissenschafte*n [The experimental method in the natural and social sciences]. Berlin, Germany: VEB.

Pedhazur, E.J., & Pedhazur-Schmelkin, L. (1991). *Measurement, design, and analysis: An integrated approach*. Hillsdale, NJ: Erlbaum.

Pfanzagl, J. (1968). *Theory of measurement*. Würzburg, Germany: Physica-Verlag.

Posner, M.I., Petersen, S.E., Fox, P.T., & Raichle, M.E. (1988). Localization of cognitive operations in the human brain. *Science*, *240*, 1627–1631.

Radder, H. (Ed.). (2003). *The philosophy of scientific experimentation*. Pittsburgh, PA: University of Pittsburgh Press.

Raichle, M.E. (1993, June 17). The scratchpad of the mind. *Nature*, *363*, 583–584.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Nielsen & Lydiche. (Original work published 1960)

Rosenthal, R., & Rosnow, R.L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.). New York: McGraw-Hill.

Scheiblechner, H. (1995). Isotonic ordinal probabilistic models (ISOP). *Psychometrika*, *60*, 281–304.

Scheiblechner, H. (1999). Additive conjoint isotonic probabilistic models (ADISOP). *Psychometrika*, *64*, 295–316.

Schönemann, P. (1994). Measurement: The reasonable ineffectiveness of mathematics in the social sciences. In I. Borg & P. Mohler (Eds.), *Trends and perspectives in empirical social research* (pp. 149–160). Berlin, Germany: Walter de Gruyter.

Scott, D., & Suppes, P. (1958). Foundational aspects of theories of measurement. *Journal of Symbolic Logic*, *23*, 113–128.

Sixtl, F. (1982). *Messmethoden der Psychologie: Theoretische Grundlagen und Probleme* [Measurement methods in psychology: Theoretical foundations and problems] (2nd ed.). Weinheim, Germany: Beltz.

Sixtl, F. (1985). Notwendigkeit und Möglichkeit einer neuen Methodenlehre der Psychologie [The necessity and possibility of a new methodology in psychology]. *Zeitschrift für experimentelle und angewandte Psychologie*, *32*, 320–339.

Sixtl, F. (1998). Der Abschied von Homme Moyen alias Average Person [Taking leave of the average person]. In W. Hacker & M. Rinck (Eds.), *Bericht über den 41. Kongreß der Deutschen Gesellschaft für Psychologie in Dresden 1998* [Report on the 41st Congress of the German Psychological Association in Dresden, 1998] (pp. 519–526). Lengerich, Germany: Pabst Science.

Stevens, S.S. (1951). Mathematics, measurement, and psychophysics. In S.S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1–49). New York: Wiley.

Suppes, P., Krantz, D.H., Luce, R.D., & Tversky, A. (1989). *Foundations of measurement: Vol. 2. Geometrical, threshold, and probabilistic representations.* New York: Academic Press.

Suppes, P., & Zinnes, J.L. (1963). Basic measurement theory. In R.D. Luce, R.R. Bush, & E.H. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 3–76). New York: Wiley.

Tarski, A. (1954). Contributions to the theory of models I. *Indagationes Mathematicae*, *16*, 572–588.

Taylor, J.R. (1997). *An introduction to error analysis* (2nd ed.). Sausalito, CA: University Science Books.

Teichmann, J. (1977). Zur Entdeckungsgeschichte des ohmschen Gesetzes, 1775–1826 [On the history of the discovery of Ohm's law, 1775–1826]. *Elektrotechnische Zeitschrift*, *98*, 96–99.

Tetens, H. (1987). *Experimentelle Erfahrung* [Experimental experience]. Hamburg, Germany: Felix Meiner Verlag.

Thurstone, L.L. (1927). A law of comparative judgment. *Psychological Review*, *34*, 273–286.

Thurstone, L.L. (1959). *The measurement of values*. Chicago: University of Chicago Press.

Trendler, G. (2006). Hugo Dingler und das Massproblem in der Psychologie [Hugo Dingler and the measurement problem in psychology]. In P. Janich (Ed.), *Wissenschaft und Leben* [Science and life] (pp. 123–135). Bielefeld, Germany: transcript.

Wigner, E. (1960). The unreasonable effectiveness of mathematics in the natural sciences. *Communications on Pure and Applied Mathematics, 13,* 1–14.

Wright, B.D. (1997). A history of social science measurement. *Educational Measurement*, *52*, 33–45.

Wright, B.D., & Linacre, J.M. (1989). Observations are always ordinal; measurements, however, must be interval. *Archives of Physical Medicine and Rehabilitation*, *70*, 857–867.

Wright, B.D., & Masters, G.N. (1982). *Rating scale analysis: Rasch measurement*. Chicago: Mesa Press.

ACKNOWLEDGEMENTS. I wish to thank Paul Barrett and Kurt Maurer for their stimulating and immensely helpful comments on different drafts of this paper. I'm especially grateful to the former for his imperturbable patience and incessant encouragement, without which this paper probably never would have been written. Valuable help was also received from Peter Janich, Joel Michell, Hans Rettler, Bernhard Orth and Hans Irtel. Last but not least I would like to express my gratitude to Henderikus J. Stam, the Editor of this journal, and to the two anonymous reviewers for their constructive comments and criticisms.

GÜNTER TRENDLER is a Ph.D. student at the Mannheim Medical Faculty of the University of Heidelberg. He is investigating the consequences of alcohol and cannabis misuse on the long-term course of schizophrenia. Previously he was involved as a graduate research assistant in the Schizophrenia Research Unit of the Central Institute of Mental Health (CIMH) in research to improve the early detection of the onset of psychosis. ADDRESS: Schizophrenia Research Group, Central Institute of Mental Health, J5, 68159 Mannheim, Germany. [email: guenter.trendler@zi-mannheim.de]