Studies in History and Philosophy of Science 42 (2011) 509–524

Contents lists available at SciVerse ScienceDirect

Studies in History and Philosophy of Science

journal homepage: www.elsevier.com/locate/shpsa

Thermoscopes, thermometers, and the foundations of measurement

David Sherry

Northern Arizona University, Flagstaff, AZ 86011-6011, United States

ARTICLE INFO

Article history:
Received 20 November 2010
Received in revised form 27 July 2011
Available online 31 August 2011

Keywords:
Measurement theory
History of thermometry
Psychological measurement

ABSTRACT

Psychologists debate whether mental attributes can be quantified or whether they admit only qualitative comparisons of more and less. Their disagreement is not merely terminological, for it bears upon the permissibility of various statistical techniques. This article contributes to the discussion in two stages. First it explains how temperature, which was originally a qualitative concept, came to occupy its position as an unquestionably quantitative concept (§§1–4). Specifically, it lays out the circumstances in which thermometers, which register quantitative (or cardinal) differences, became distinguishable from thermoscopes, which register merely qualitative (or ordinal) differences. I argue that this distinction became possible thanks to the work of Joseph Black, ca. 1760. Second, the article contends that the model implicit in temperature's quantitative status offers a better way for thinking about the quantitative status of mental attributes than models from measurement theory (§§5–6).

1. Introduction

Psychologists debate whether mental attributes can be quantified or whether they admit only qualitative comparisons of more and less. Their disagreement is not merely terminological, for it bears upon the permissibility of various statistical techniques (Cp. Stevens, 1951, 25ff.; Lord, 1953). This article contributes to the discussion in two stages. First it explains how temperature, which was originally a qualitative concept, came to occupy its position as an unquestionably quantitative concept (§§1–4). Specifically, it lays out the circumstances in which thermometers, which register quantitative (or cardinal) differences, became distinguishable from thermoscopes, which register merely qualitative (or ordinal) differences. I argue that this distinction became possible thanks to the work of Joseph Black, ca. 1760. Second, the article contends that the model implicit in temperature's quantitative status offers a better way for thinking about the quantitative status of mental attributes than models from measurement theory (§§5–6).

Mental attributes, like intelligence and aggressiveness, have degrees that can be ordered. Psychologists have taken these ordinal characteristics as a sign that mental attributes are quantitative. Michell, a critic of psychological measurement, thinks this inference is presumptuous (1990, 170), and he contends that the history of the temperature concept is partly to blame for the presumption: "[T]he fact that some physical quantities were initially identified only ordinally (e.g., temperature) has encouraged psychologists to treat order as a sign of quantity" (ibid.). Order doesn't entail quantity, of course, but why shouldn't psychologists find encouragement in the success of thermometry? Michell's response would seem to be that thermometry's success counts for little now that there exists a body of theory, which enables psychologists to test hypotheses that mental attributes have quantitative structure (1999, 219). For, according to Michell,

> [I]n the absence of experimental tests known to be specifically sensitive to the hypothesised additive structure of the attribute studied, it is not known whether or not these attributes are quantitative and thus it is not known whether or not existing procedures measure them. (1999, 216)

It is with advances in measurement theory, in particular, conjoint measurement, that tests sensitive to additive structure in intensive magnitudes have become possible. I argue that such tests are less valuable than simply looking for benefits analogous to those that Black derived from treating temperature as a quantity.

E-mail address: David.Sherry@nau.edu

0039-3681/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.shpsa.2011.07.001

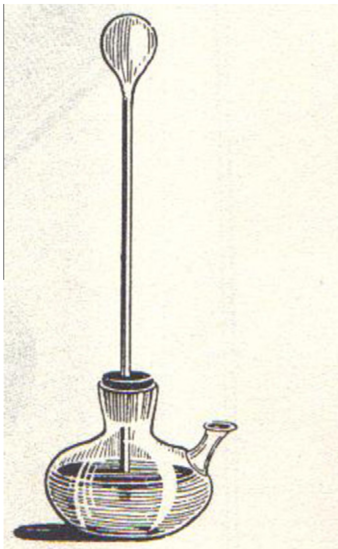

Fig. 1. Galileo's thermoscope.

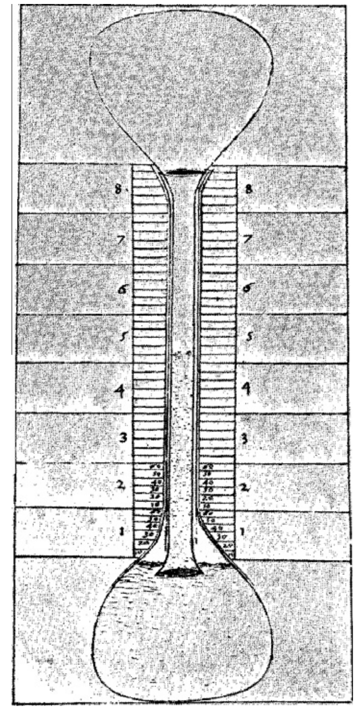

Fig. 2. Telioux's thermoscope-cum-scale.

2. Thermoscopes and thermometers

Galen recognized four degrees of heat and four degrees of cold, and classified drugs according to their power to heat or cool a patient to a specified degree (Taylor, 1942, 129–130). However, Galen's use of degrees depended only upon his judgment as a physician, not the use of a measurement procedure. He did not count degrees, but only compared his patients to standard cases. His procedure is similar to judging that a patient has first, second or third degree burns: A suitable series of adjectives (moderate, serious, severe, extreme, etc.) would serve as well. Thus, Galen's use of ''degree'' did not *require* him to assign a numeral. For present purposes, ''temperature'' will denote a degree (or level) of heat *independent* of its being assigned a numeral. Our goal is to unearth the circumstances in which assigning a numeral to a temperature constitutes a quantitative measurement.

Thermoscopes were the earliest devices for detecting temperature. Historians generally credit Galileo with constructing the first one, ca. 1592. His thermoscope worked on the principle—known in antiquity—that air expands when heated. When Galileo heated the air in his thermoscope (Fig. 1), air was driven out of the tube and bubbled up through the water. When the air cooled, water would rise in the tube. Galileo's device is affected by air pressure, so, strictly speaking, it is a barothermoscope. When air pressure is not a factor, a thermoscope enables one to observe sameness and difference in temperature.

For some authors, a thermometer is just a thermoscope with a scale (e.g., Middleton, 1966, 4). But what is a scale? And will any sort of scale suffice for quantitative measurement? Middleton recognizes a drawing and description of a thermoscope-cum-scale from 1611 (Fig. 2). The drawing and description are by Telioux. While Middleton treats this instrument as an air thermometer, it's less sophisticated an instrument than that constructed by von Guericke, ca. 1660 (Fig. 3). Von Guericke's device is unaffected by air pressure because it is closed to the atmosphere, and the 'working liquid' is brandy, which doesn't freeze at colder temperatures. Like Telioux's instrument, von Guericke's gives readings in one of eight regions. But *unlike* Telioux's instrument, von Guericke's makes no pretense of measurement. Yet it gives more and better information than Telioux's, not only because air pressure is not a factor, but because ''frigid air,'' ''temperate air,'' etc., which correspond to familiar sensations, are *less* arbitrary than Telioux's numerals.

In view of the last two examples, it is unsurprising that there is little consensus about the appearance of the first thermometer. Campbell, who interprets ''measurement'' quite strictly, writes:

> Thus, the scale of temperature of the mercury in glass Centigrade thermometer is quite as arbitrary as that of the instrument with the random marks. (1920, 359)
> The international scale of temperature is as arbitrary as Mohs' scale of hardness. (400)[1]

Other authors are less extremes. Wolf, for example, refers indifferently to Galileo's device and its immediate descendants with both ''air-thermometer'' and ''air-thermoscope'' (1950, 83ff.); though once the discussion turns to liquid devices, he drops the term ''thermoscope'' (86ff.). Wolf later distinguishes thermometers from thermoscopes by the presence of a scale, though he blurs his distinction by observing that at first these degrees were ''purely arbitrary in value.'' *This* problem, he claims, was solved at the end of the 17th century by the adoption of two fixed points ''and the division of the interval between these points into a conventional number of equal parts'' (1952, 306–307).[2] Unfortunately, he never clarifies why *that* should be the decisive event.

Roller links the first thermometer to a scale that could be reproduced with precision (1966, 121–123). Thus, he treats the instruments of Accademia del Cimento (ca. 1660) as thermoscopes because their fixed points, most severe winter cold and greatest

[1] Campbell is referring not only to Fahrenheit's 18th century instruments, but to the international temperature scale adopted by the Fifth General Conference of Weights and Measures in 1913. See ''Report of the National Physical Laboratory for the Year 1928,'' 29–33, http://www.rsc.org/delivery/_ArticleLinking/DisplayArticleForFree.cfm?doi=AN9295400292&JournalCode=AN.

[2] Perhaps Wolf sees Renaldini's spirit thermometer, which introduced the ice and steam points as fixed points ca. 1694, as the first *real* thermometer. But he argues elsewhere that De Luc demonstrated the validity of the conventional division into equal parts only in the second half of the 18th century and only for the mercury thermometer (294).

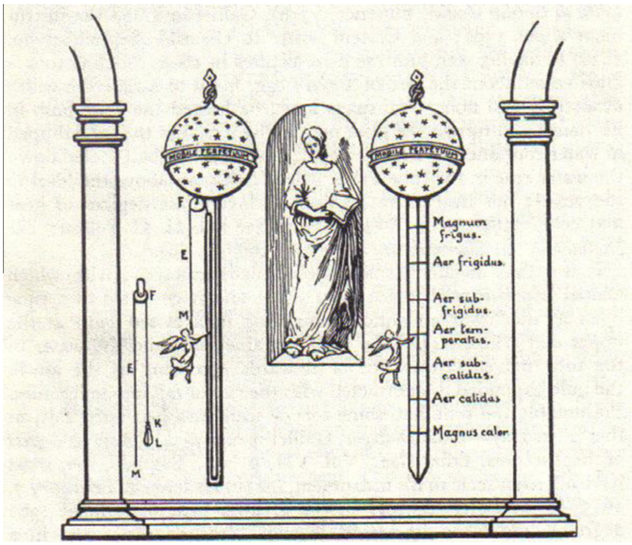

Fig. 3. von Guericke's thermoscope/thermometer.

summer heat, could not be determined with precision. He places the emergence of the thermometer at the point of recognizing that air pressure must be taken into consideration before the ice and steam points can be suitably fixed to constitute a universally comparable scale. Roller, too, offers no justification for his decision to found quantitative measurement upon universal comparability.

Klein contends that the first thermometer is probably due to Rømer (ca. 1702), who introduced a method for determining rigorously the uniformity of the tube through which liquid rose and fell (1974, 296ff; cf. Middleton, 1966, 67). This division point is also prima facie plausible. But like Wolf's and Roller's it employs a concept of quantitative measurement without making it explicit.

A recent and generally fascinating history of thermometry, Chang (2004), also fails to explicate quantitative measurement. For instance, in the course of arguing that there is more to a scale than simply making a series of marks on a tube, Chang writes

> To help our observations, some lines can be etched onto the tube, and some arbitrary numbers may be attached to the lines. Many of the early instruments were in fact of this primitive type. These instruments should be carefully distinguished from thermometers as we know them, since they are not graduated by any principles that would give systematic meaning to their readings even when they are ostensibly quantitative. I will follow Middleton in dubbing such qualitative instruments *thermoscopes*, reserving the term *thermometer* for instruments with quantitative scales that are determined on some identifiable method. (Chang, 2004, 41)

Chang's condition, "graduated by ... principles that would give systematic meaning to their readings," does little to reveal the principle that distinguishes thermoscopes from thermometers. For some (Wolf, e.g.) are evidently satisfied that Fahrenheit (ca. 1720) gave systematic meaning to readings from his device, while others (Campbell, e.g.) would just as evidently deny that claim. Subsequently, Chang hints that systematic meaning requires "numbers for which some arithmetical operations yield meaningful results" (ibid.). Later, however, Chang confounds the issue by distinguishing ordinal quantities from cardinal quantities (229). If there are ordinal *quantities*, then one can't "reserve the term thermometer for instruments with quantitative scales that are determined on some identifiable method," for thermoscopes use ordinal quantitative scales.

Chang's gesture toward meaningful arithmetical operations offers a promising strategy, but he doesn't pursue it. Instead, after drawing the line between thermoscopes and thermometers at the introduction of water's freezing and boiling points, he simply declares that this innovation

> ... allowed a true quantification of temperature. By means of numerical thermometers, meaningful calculations involving temperature and heat could be made and thermometric observations became possible subjects for mathematical theorizing. (48)

Even if this account is correct, it lacks an argument to show that finding phenomena sufficiently constant in temperature (as determined by thermoscopes) to serve as fixed points, and then dividing the interval between them *suffices* to quantify temperature. We need, at a minimum, an account of meaningful calculations involving temperature and heat. §§3–4 provide such an account, based upon the different *uses* of thermoscopes and thermometers.

3. Ranking

Letters to Galileo from his friend Sagredo include many examples of the early uses of thermoscopes. In 1613 Sagredo wrote:

> The instrument for measuring heat, invented by your excellent self, has been reduced by me to various very elegant and convenient forms, so that the difference in temperature between one room and another is seen to be as much as 100 degrees. With these I have found various marvelous things, as, for example, that in winter that air may be colder than ice or snow; that the water just now appears colder than the air; that small bodies of water are colder than large ones, and similar subtle matters. (quoted in Middleton, 1966, 6–7)

Of Sagredo's four examples, the last three involve a comparison by means of the relation 'hotter than' between one body and another. His judgments stem from observing that the bodies correspond to different levels in a thermoscope; none of these observations require that a numerical scale be attached to the instrument. However, the first example purports to observe a quantitative difference between the levels of the thermoscope, and that requires a numerical scale. Do the numerals on this scale serve a purpose beyond, say, that of "frigid air," "temperate air," etc. in von Guericke's scale?

Sagredo's scale includes infinitely more degrees of temperature; so, it yields a *more precise* ranking of thermal phenomena. It is also more convenient than von Guericke's. Von Guericke requires separate procedures to locate each of the intermediate points, while Sagredo's intermediate points are established by simple geometry. There is, however, no indication that Sagredo put observations of differences and ratios of differences to use by correlating them with further phenomena. Had he been able to observe that, for example, equal temperature differences correlate mathematically with differences between magnitudes of some further property—as Galileo was able to correlate differences between time consumed and distances traversed in free fall—he would have laid claim to practical application of his instrument. But Sagredo was mostly content to marvel at rankings, e.g., well water's being colder in winter than summer, in spite of what the senses reveal (7).[3]

[3] Sagredo also reported proportions between differences in temperature. "[I]t appears that salt combined with snow increases the cold by as much as amounts to a third of the difference between the excessive heat of summer and the excessive cold of winter" (quoted in Middleton, 1966, 10).

Kuhn comments that early experiments in the study of heat are more like investigations *of* a new instrument rather than investigations *with* it (1961, 58). That seems exactly right: Sagredo's 'quantitative' observations suggest a tremendous potential for his instrument, but he was unable to actualize that potential. Kuhn, in fact, was one of the first to appreciate that a body of theory is ordinarily pre-requisite for fruitful measurement (ibid., 47). In the absence of some effort to coordinate phenomena by means of *mathematical* laws, Sagredo's measurements could do little more than nourish curiosity about the realm of thermal phenomena.

Sanctorius *did* do more than investigate a new instrument. He took the temperature of patients by observing the distance through which the liquid fell during ten beats of a small pendulum (Taylor, 1942, 138–139). When the liquid fell more quickly, it was an indication of fever.[4] Sanctorius's observations, then, had a use beyond merely ranking thermal phenomena. They were used to draw further inferences: The symptoms associated with fevered or non-fevered states could be inferred from the rate at which the column of liquid fell. And perhaps increasingly severe symptoms could be inferred from increasingly rapid rates of fall. Even so, Sanctorius didn't require that the thermal levels be quantitative. His purpose can be accomplished by observing merely qualitative relations: He ranked thermal phenomena by ranking distances. Sanctorius's inferences are not drawn in accordance with arithmetic; for they depend only upon the order of the numerals.

None of this suggests that thermoscopes had no role in physical theory. Boyle's law (1662) states that for a sample of a gas at a constant temperature, as the volume increases the pressure decreases proportionally, and vice versa. In order to demonstrate this law experimentally, Boyle had to subject readings of volume and pressure to arithmetic treatment; he showed that *for a given temperature* the products of correlated measures of volume and pressure gave were equal to one another. This result depends upon the level indicated by a thermoscope. Boyle had observed that temperature increased with pressure, and so he had to apply a cold cloth to the bulb containing the gas in order to bring the temperature down (Schooley, 1986, 12). But even if the levels on Boyle's thermoscope were marked by numerals, he did not employ these numerals in calculation. That is, his demonstration of the law did not require that temperature be a genuine quantity. Likewise, Boyle's determination of the expansion coefficients for different liquids did not presuppose that temperature is a quantitative attribute. For he would measure the degree of contraction or expansion between fixed points, e.g., room temperature and freezing (Barnett, 1956, 290). And although he and his successors often employed Boyle's law, that employment (as far as I have been able to determine) never presupposed that temperature possessed more than an ordinal structure. That is, none of these advances achieved a theoretical understanding of thermal phenomena by subjecting temperature readings to arithmetic operations.

There were, however, a series of technical achievements that had to occur before the quantification of temperature was even possible. They solved a problem articulated by Huygens in a letter of 1665:

> It would be a good thing to devise a universal and determinate measure of cold and heat ... so that without sending thermometers[5] one should be able to communicate the degrees of heat and cold which have been found in experiments, and to record these for posterity. (quoted in Middleton, 1966, 50–51)

Reliable instruments must give the same readings in the same circumstances. Their method of production should be such that they achieve the effect of a single instrument *without* the inconvenience of sharing. Reliable instruments are, in other words, *comparable*, and comparability is the criterion to which Wolf, Roller, and Chang appeal in dating the appearance of thermometers.

The problem of comparability was solved gradually, as instrument makers became aware of the conditions under which readings could vary. It's understandable, then, that historians could disagree about the date of the thermometer's appearance. In the 1640s scientists[6] began to understand that thermoscopes were affected by atmospheric pressure. To solve this problem instrument makers used liquid instead of air as the thermometric substance, i.e., the substance whose states are the observed effects of temperature, and sealed off the thermometric substance from the atmosphere (Middleton, 1966, 27–28). The next steps toward comparability were settling upon one thermometric substance and one method for numbering points along the scale. By 1740—shortly after Fahrenheit's death and as Celsius was beginning to make instruments—mercury had become the preferred substance: It is easy to expel impurities from mercury; it remains liquid over a wide range (−38 °F to 600 °F); unlike water and alcohol, it does not cling to the sides of the tube; and, unlike water, it does not expand as it approaches the lower limit of fluidity. At the end of the 17th century, the preferred method of numbering the scale was to choose a pair of fixed points and to define a degree as a fixed proportion of that interval. By 1740 instrument makers had settled upon the freezing and boiling points of water, and as a result of these and other innovations 'thermometers'[7] had become fairly well standardized. For example, a pair of Fahrenheit's devices read within a sixteenth of a degree of one another (Middleton, 1966, 74). Still, no more could be claimed for them than that they provided standards for a uniform and precise *ranking* of thermal phenomena. There were still no meaningful calculations based on readings from Fahrenheit's instruments.

4. Quantifying

Joseph Black is the first scientist who can lay claim to having employed Fahrenheit's device to quantify intensity of heat. He accomplished this feat in the 1760s, decades after Fahrenheit's death. Unlike his predecessors, Black used readings from his device to achieve a *theoretical* understanding of thermal phenomena by *applying mathematics* to the readings. Surprisingly, Black's mathematical techniques emerged long before the thermoscope. He melded this older, conceptual tradition with the experimental tradition that immediately preceded his discoveries.

4.1. Conceptual foundations

The conceptual tradition arose in the late Middle Ages, reaching its pinnacle in the work of Nicole Oresme, ca. 1350. While Oresme and his colleagues investigated topics that can be traced to Aristotle, their results were a marked departure from his thought. Aristotle distinguished ultimate categories of being, among them

[4] The liquid *falls* in the presence of fever because Sanctorius used an air thermoscope.

[5] See note 7 below.

[6] "Natural philosopher" is perhaps more suitable, as "scientist" did not appear in English until the 1760's.

[7] The term *thermoscope* (*thermoscopium*) appeared first in 1617 (Middleton, 1966, 11). Only seven years later the term *thermometer* (*thermomètre*) was used first by Leurechon (Wolf, 1950, 84). In the seventeenth century the term "thermometer" was applied to instruments whose readings were put to no strictly quantitative use. See the quote from Huygens above and Taylor (1942, 132). The distinction between thermoscopes and thermometers is, therefore, an expository device. That device becomes problematic in the context of Fahrenheit's 'thermometer', which could claim the mantle of thermometer only decades after Fahrenheit's death, once Black used it to construct new thermodynamic concepts.

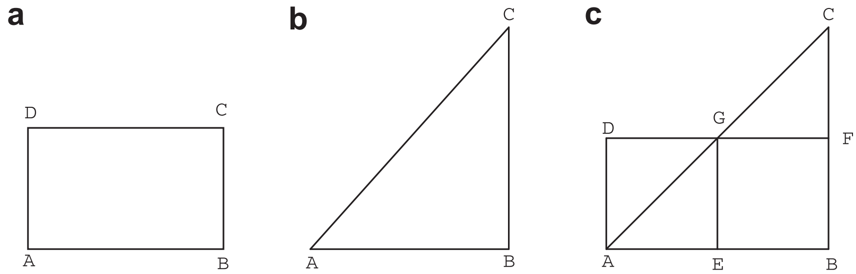

Fig. 4. Oresme's diagrams.

quality and quantity (***Cat.*** 1b25ff.). Some of the former (e.g., white, grammatical) admit of contraries as well as more and less (10b12, 10b26), while none of the latter (e.g., four-foot, five-foot) admit contrariety or more and less (5b11, 6a19).[8] Although Aristotle recognized different degrees of quality (***Phys.*** 226b1-5), he denied the possibility of comparing them quantitatively.

> That which is not a quantity can by no means, it would seem, be termed equal or unequal to anything else. One particular disposition or one particular quality, such as whiteness, is by no means compared with another in terms of equality and inequality but rather in terms of *similarity*. (6a31-4, my italics)

Even if Aristotle were willing to compare differences between degrees, there would have been no question of comparing those differences in terms of equality and inequality.

Aristotle did not always adhere to the strict separation of quality from quantity. For instance, he treated "thin" and "fast" as standing in quantitative relations like double and half (***Phys.*** 215b1-11). Furthermore, he considered a quantitative treatment of degrees of hot and cold when he suggested that a mixture of hot and cold elements might have a "power-of-heating that is double or triple its power-of cooling" (***De Gen. & Corr.*** 334b8-16). Aristotle, it seems, succumbed to the temptation to take order as a sign of quantity. Medieval physicists were similarly tempted, and much to the benefit of posterity. For in developing accounts of qualitative change they set aside Aristotle's rigorous distinction between quality and quantity and treated qualities as continuous quantities.[9] The resulting conceptual apparatus made Black's achievements possible.

Yet the medievals did not ignore entirely Aristotle's distinction, for they referred to the degrees of quality as *intensive* (or *virtual*) quantities, distinguishing them from *extensive* (*dimensive* or *corporeal*) quantities, which are quantities in the sense of Aristotle's ***Categories*** (Clagett, 1959, 212). According to Aristotle, quantities are always composed of parts (4b20-1); intensive quantities, in contrast, lack parts, or at least parts that are apparent. Medieval physicists were able to advance the quantitative treatment of qualities by turning attention to the distribution of a quality in a subject. Intensity of heat can be distributed uniformly or non-uniformly in a body; for example, the intensity of heat may be less in exposed than unexposed flesh. Likewise, intensity of motion (velocity) can be distributed uniformly or non-uniformly over the interval in which a motion occurs, as when a body moves at a constant velocity or accelerates. When a quantitative treatment of qualitative intensity was applied to the analysis of variation in intensity, a new field of research arose, viz., the quantitative study of the extension of a quality in a subject. Although Oresme was not the first to contribute to this field, his geometric treatment of the subject clarified and extended it significantly (cf. Oresme, 1968).

Oresme proposed that the extension of a quality in a subject be represented by a figure whose base corresponds to the extension of the subject and whose altitudes correspond to the intensity of that quality at each point in the subject. Thus, a rectangle represents a quality whose intensity is uniform throughout its subject (Fig. 4a), while a right triangle (or right trapezoid) represents a quality whose intensity is uniformly non-uniform (or uniformly difform) (Fig. 4b).

The areas of Oresme's figures constitute measures of the *quantity* of a quality, and by comparing these areas, it is possible to demonstrate theorems about the quantity of a quality. One, the mean degree theorem, reappeared 300 years hence in Galileo's demonstration that distance in free fall is proportional to the square of the time (Galilei, 1974, 165–169). The mean degree theorem states (cf. Fig. 4c): The quantity of a uniformly difform quality (ΔABC) is equivalent to the quantity of quality uniform in the degree mean between the initial and final degrees of the latitude uniformly difform (ABFD, with BF = 1/2 BC). The mean degree theorem treats both intensities and differences between intensities as continuous quantities. Oresme's graphical representations are especially perspicuous, but arithmetic and algebraic representations of the quantity of a quality were employed both before and after Oresme. In such cases the quantity of a quality was understood as the product of an intension and an extension. Newtonian mechanics made particularly good use of this technique, e.g. mass = density (intension) $\times$ volume (extension) (Roche, 1998, 117–118). Black's conceptual innovations—specific and latent heats—used the same strategy.

By representing quantities of heat as the product of intensions (i.e., temperatures) and extensions, it is possible to solve problems such as the following mixture problem from a treatise that has been attributed to Roger Bacon (late 13th century).

> [L]et there be given water of two weights hot in the sixth degree, ... [and] let there be given again another water of one weight hot in the twelfth degree; ... a mixture of the two waters having been made, the hotness of the mixture will be raised in a line intension through eight degrees, since the distance that is between six and eight is one half the distance that is between eight and twelve, just as the water of one weight is half the water of two weights. (Clagett, 1959, 335)

The author imagines that the final temperature (t_f) achieved by both liquids results from two pounds at six degrees gaining a quantity of

[8] As a consequence, length is not a quantity, though particular lengths (four-foot, five-foot) are. Thus Aristotle can assert, "Quantity does not, it appears, admit of variation of degree" (***Cat.*** 6a19), a claim that is patently false in modern measurement theory. There length is a paradigm quantity in which Aristotle's quantities (four-foot, five-foot) – are *levels* or *degrees*.

[9] Murdoch attributes the change to the emergence, poorly understood, of "a veritable furor to measure all things possible" (1974, 62). The furor is amply illustrated in Crosby (1997), which, unfortunately, doesn't explain the furor either. In any case, the furor was to quantify, rather than measure; medieval physicists never actually measured anything. Maier (1982) suggests that the urge to quantify can be traced to scripture (149–150): "You have ordered all things in measure, number and weight" (Wisdom 11:21).

heat that is lost by one pound at twelve degrees; in other words, 2 lbs. × (t_f − 6) = 1 lb. × (12 − t_f), and so t_f = 8 degrees. Call this "Bacon's rule." The rule rests upon a principle for computing weighted means, here, $t_f = \frac{w_1 t_1 + w_2 t_2}{w_1 + w_2}$. There is nothing mathematically new in the mixture problem, as ancient computations of balance points used an analogous principle. The only innovation consists in treating a quality (temperature) as a continuous quantity. Of course the medieval investigation of quantities of quality proceeded *secundum imaginationem*—on the supposition that such measurements had been made—as no thought was given to actual measurement. But four centuries later these flights of imagination provided the means for introducing a quantitative temperature concept.

4.2. Non-arbitrariness

In order for Fahrenheit's device to measure quantity, its scale must not be, in Campbell's phrase, "as arbitrary as Mohs' scale of hardness" (1920, 400; cf. p. 5 above). The arbitrariness of Mohs' scale lies in the freedom available to Mohs in assigning numerals to different levels of hardness. The only restriction on his assignments is that the numeral assigned to the harder mineral come *after* the numeral assigned to the softer mineral, given some procedure for determining which of the two minerals is harder than the other. That is, any monotonic transformation of Mohs' scale would have served Mohs' purpose—viz., ranking—equally well. Relative to Mohs' scale the difference between the hardness of diamond (10) and the hardness of quartz (7) is 3, and likewise the difference between the hardness of quartz and the hardness of fluorite (4) is 3. But one may not infer that quartz is harder than diamond by an amount equal to that by which diamond is harder than quartz *unless* any monotonic transformation of Mohs' scale would dictate the same inference. Obviously, that condition fails to hold; the transformation f: $n \to 2^n$, for instance, would not yield an equality.

Early developers of thermometers no doubt *hoped* that their instruments would do more than rank thermal phenomena. Boyle, for instance, wrote in 1662,

> I consider that we are very much to seek for a standard, or certain measure of cold, as we have settled standards for weight and magnitude and time; so that when a man mentions an acre, or an ounce, or an hour, they that hear him know what he means and can easily exhibit the same measure. (quoted in Barnett, 1956, 304)

But even the technical innovations of Fahrenheit et al. could not guarantee that the equally spaced subintervals on a mercury thermometer are more than a precise and convenient way to compare different levels of temperature. Their equal subintervals would have lack significance if, for instance, mercury expanded slowly at temperatures below blood heat but more rapidly at temperatures above blood heat.[10]

To rule out such a scenario would be easy if there were a way of determining how mercury expands with the rise of temperature, i.e., if one had a function from temperature to volume, $f(temp_{Hg}) = vol_{Hg}$. But such a function presupposes a thermometer whose non-arbitrariness is already established. This vicious circle, which seems to dash the hope of a non-arbitrary temperature scale, is not always appreciated. Black observed that his predecessor Boerhaave overlooked the question (Black, 1803, 57). McGee (1988), a standard text on temperature measurement, claims that Fahrenheit switched from alcohol to mercury "because it had a more nearly linear thermal expansion with temperature" (7). Yet there is no indication of how Fahrenheit could have known this.

Chang refers to the vicious circle as "the problem of nomic measurement" (2004, 59–60). He first noticed the problem in connection with energy measurements in quantum mechanics, where he settles for a tolerable circularity, allowing different methods of measurement to justify one another (1995, 153–154). He believes, however, that the problem of nomic measurement is solvable for temperature.

> ... Regnault solved the greatest problem standing in the way of making numerical temperature observable: the problem of nomic measurement. The solution, as I have already noted, was the criterion of comparability ... (2004, 89).

But Regnault's account of comparability is not up to solving the problem of nomic measurement:

> Regnault's secret was the idea of "comparability." If a thermometer is to give us the true temperatures, it must at least always give us the same reading under the same circumstance. ... Regnault considered this "an essential condition that all measuring apparatuses must satisfy." (77)

While comparability is necessary for an instrument to *rank* temperatures for scientific ends, the fact that an instrument gives the same readings under same circumstances is no indication that a given attribute is quantitative. As observed a few paragraphs back, comparability can't insure equality of differences given the possibility that the thermometric substance expands differently in different temperature ranges. Consider the following analogous case. It is not out of the question that with practice a panel of judges could achieve uniformity in assigning scores.[11] But it would not follow that the judges' agreement shows that they are measuring a continuous quantity; for nothing in this scenario commits them to claiming that the difference between samples marked 5 and 3 is equal to the difference between essays marked 10 and 8. Thus nomic measurement is still, apparently, a difficulty for temperature.

Michell, too, is casual about the problem of nomic measurement. In describing how scientists carry out the instrumental task of quantifying temperature he writes,

> Within a specific range of temperatures, it has been found that the temperature of a liquid (say, the metal, mercury) is linearly related to its volume, if pressure is held constant. Thus in a sealed glass tube of uniform width, for a limited range, temperature varies linearly with the height of the column of liquid. By this mean it is possible to measure temperature via measurements of length. (1999, 76)

But nowhere does Michell present the warrant for the claim that, within some range, mercury's temperature is linearly related to its volume, although the claim presupposes that one has already a method of measuring temperature. Since this case has, apparently, misled psychologists, it is incumbent upon Michell to elaborate. In my view, the elaboration would feature the Scottish chemist Black, the first to overcome temperature's problem with nomic measurement.

Black sought to understand quantitatively how heat flowed in processes like melting, boiling, and mixing substances at different temperatures. He would not have undertaken a quantitative investigation of heat flow unless he believed, with his predecessors, that temperature is a continuous quantity, similar to length, time, and weight. Length, time, and weight have parts, as Aristotle observed, and because those parts may be joined or separated, it is apparent that measures of length, time, and weight may be added to and

[10] Using mercury as a standard, De Luc (1772) found this is exactly what happens with alcohol and other thermometric substances (Middleton, 1966, 124).

[11] Uniformity is important in studies on the effectiveness of treatments, and a great deal of time and money are spent on insuring uniform judgments, even when the data is non-numerical.

subtracted from one another to calculate the result of joining and separating. But temperature lacks parts; thus it is not apparent that thermometer readings may be added or subtracted. Black's results vindicated his belief that thermometer readings may be added and subtracted.

In order to study heat flow, Black needed an instrument whose readings could be subject to calculation; so he acquired one of Fahrenheit's thermometers and tested whether its readings satisfied Bacon's rule (Black, 1803, I, 56–59).[12] If the device *really* measured temperature, then its readings of final temperatures of mixtures should agree with temperatures calculated using that rule. For instance, if Fahrenheit's instrument reads 100° inserted in a sample of water weighing 1 pound and 50° in another sample of water weighing two pounds, then if the device is measuring a continuous quantity, the final temperature should read $\frac{1(100)+2(50)}{1+2} = 66.6$. By varying weights and temperatures, it should be possible to check Fahrenheit's device at most points on the scale. Black writes:

> The experiment being made with these precautions, the result has shown, that when the thermometer is made of quicksilver, the gradual expansions of this fluid, while it is heated slowly, from the cold of melting snow to the heat of boiling water are very nearly proportional to the additions of heat by which they are produced. There is, however, a little deviation from the exact proportion; while the heat increases, the expansions become a little greater than in proportion to the increase or addition of heat. . . . In [quicksilver] this irregularity is so inconsiderable within the range of heat above mentioned, that it does not deserve any notice in common experiments, and in the ordinary use of thermometers. (59)

Evidently, he was satisfied.

Black's test is standard hypothetico-deductive procedure, viz., deducing future readings from prior readings and the hypothesis that the instrument's readings obey the laws of continuous quantity. Black's result is grounds to hope that Fahrenheit's device sheds light on heat flow. But, as one expects from a hypothetico-deductive test, it cannot verify that "the degrees of their scales express, or point out, equal differences of heat" (56). There are at least three difficulties that arise. (1) Black doesn't give specific predictions, but when DeLuc performed the experiment in 1772, he found that mixing equal weights of water at 0° and 80° (on Réaumur's scale) yielded 38.6° rather than the predicted 40°. Whether to count this as a confirmatory or falsifying instance depends on the value of the uses to which the readings are put. At the time Black performed the experiments, however, temperature had not yet come to occupy a significant quantitative role, either *practical* or *theoretical*. In the absence of a useful practical or theoretical role for readings from the thermometer, the results of Black's mixture experiments do no more than reveal a curious quantitative phenomenon. I.e., the readings have yet no place in quantifying thermal phenomena.

(2) One of the precautions that Black mentions was compensating for the temperature of the vessel in which the mixing took place by performing the experiment in identical vessels, one hot and one cold, and then "taking the medium of the results" (58). Black's proposal is clever, but it's not a complete solution of the difficulty. One needs to understand the thermal behavior of materials other than mercury in order to assess the significance of the experimental results, and that, of course, is the theory Black is trying to construct. It is particularly important to understand how glass expands upon heating before one can be confident about Fahrenheit's device.[13] (3) And finally, there are unstated theoretical assumptions at the basis of Black's experiment, the conservation of heat, for example. A genuine thermometer would be useful in confirming them, too. Black too faced the problem of nomic measurement: To verify rigorously that Fahrenheit's degrees represent equal differences, he apparently required a thermometer with established credentials.

Black treated Fahrenheit's instrument as a *bona fide* measurement device, even though one could object reasonably to his doing so. Our three objections would have been decisive if Black hadn't been able to *use* the instrument in company with the medieval quantification of quantity, to turn familiar qualitative observations into a pair of quantitative concepts fundamental to the science of heat—capacity for heat[14] (nowadays, specific heat) and latent heat.

4.3. Capacity for heat and latent heat

Black's predecessors observed that when iron and wood are heated to the same temperature (indicated by a thermoscope), the iron feels warmer and feels warm longer than the wood. Likewise, they observed that when iron and wood are cooled to the same temperature, the iron feels colder and feels cold longer than the wood. Black explained these phenomena by supposing that iron had a greater capacity for heat than wood; thus iron released more heat when it was warmer and absorbed more heat when it was cooler (1803, 78). The same supposition explained the surprising observation that water took more time to become hotter and also more time to become cooler than mercury, in spite of the latter's greater density (82). Thus far, Black's explanation is merely qualitative; the heat capacities of substances can be ranked in indefinitely many levels by comparing periods of time. Black then proceeded to quantify heat capacity by means of Fahrenheit's device:

> [I] estimated the capacities, by mixing the two bodies in equal masses, but of different temperatures; and then stated their capacities as inversely proportional to the changes of temperature of each by the mixture. Thus, a pound of gold, of the temperature 150°, being suddenly mixed with a pound of water, of the temperature of 50°, raises it to 55° nearly: Therefore the capacity of gold is to that of an equal weight of water as 5 to 95 or as 1 to 19; for the gold loses 95° and the water gains 5°. (506)

Black's reasoning here is perfectly analogous to the medieval mixture experiment. In that case, temperature change was inversely proportional to weight, $\frac{\text{lighter}}{\text{heavier}} : \frac{\Delta t_{\text{heavier}}}{\Delta t_{\text{lighter}}}$; here, analogously, $\frac{\text{heat capacity}_1}{\text{heat capacity}_2} : \frac{\Delta t_2}{\Delta t_1}$. Having quantified heat capacity, Black was able to expand the scope of mixture problems to mixtures of *different* substances. Two substances of masses m_1 and m_2, capacities for heat c_1 and c_2, and initial temperatures t_1 and t_2 (with $t_1 > t_2$, say) will, upon mixing, attain the final temperature t_f satisfying the equation $m_1c_1(t_1 - t_f) = m_2c_2(t_f - t_2)$. Heat capacities as determined by Fahrenheit's device not only brought mixture problems from the conceptual realm to the empirical realm, but they managed to unify a range of phenomena much wider than that envisioned by the medieval thinkers who first *imagined* that heat phenomena could be treated quantitatively. Results were not perfect, of course, as they

[12] Black died in 1799; the text was prepared from Black's notes by his student and then colleague Robison. According to Robison, Black's experiments were carried out between 1759 and 1763 (McKie & Heathcote, 1935, 31).

[13] DeLuc, like Black before him, had found that mercury seems to expand more quickly as the temperature rises. Yet when compared with Kelvin's thermodynamic scale, the foundation of modern thermometry, the mercury thermometer actually expand more quickly at *lower* temperatures (Middleton, 1966, 125).

[14] In modern physics capacity for heat is the product of mass and specific heat. The latter corresponds to Black's capacity for heat, except that the modern concept is an absolute measure (J/kg·K), whereas Black's concept is relative measure—the ratio between the heats required to raise a given mass of a substance and an equivalent mass of water 1 °F.

depended on careful determinations of heat capacity as well as the care in taking account of heat flow to the atmosphere and the mixing vessels. But Black's construction of a quantitative concept of heat capacity launched the science of calorimetry.

$m_1c_1(t_1 - t_f) = m_2c_2(t_f - t_2)$ is the founding principle of calorimetry. It presupposes that temperature is a quantitative attribute; that is, the principle lacks sense *unless* temperature is a quantitative attribute. Black himself presumed, on the basis of his mixture experiments, that Fahrenheit's thermometer gave a good estimate of that quantity and, consequently, the quantity he called heat capacity. Within the range of temperatures to which he applied the Fahrenheit thermometer Black's presumption was supported by observations. There was no guarantee that the readings would be similarly useful outside this limited range, and indeed later comparisons of air and mercury thermometers by Dulong and Petit (1817) cast doubt on the mercury thermometer for temperatures above 300 °C (Barnett, 1956, 323–324). But doubts about the mercury thermometer didn't undermine the quantitative status of temperature, once Black had demonstrated that, within a limited range, heat capacity is a valuable concept for explaining and predicting thermal phenomena. The very possibility of using heat capacity as a theoretical concept requires that temperature be a quantitative attribute. This situation illustrates what Toulmin calls the stratification of physical theory: Statements at one level have a meaning only within the scope of those at the level below (Toulmin, 1960, 80–81). In this instance, statements (and questions) about heat capacity are unintelligible outside the scope of statements about ratios of quantitative differences between temperatures. Had Black's ideas not panned out, the point about stratification would carry no weight. But we cannot simultaneously grant that Black advanced our understanding of thermal phenomena and entertain doubts that temperature is a continuous quantity. The same argument can be made about latent heat, Black's subsequent advance in the study of heat.

The success of Black's quantification of heat capacity demonstrated that temperature change could be used as a measure of heat change (ΔQ), in accordance with the simple formula $\Delta Q = mc\Delta t$. Black was able to apply his technique for measuring heat flow in quantifying a further concept viz., latent heat. Like heat capacity, latent heat also arose from reflection upon familiar phenomena. Black's predecessors assumed that once a solid had reached its melting point, only a small addition of heat was necessary to change the entire mass to a liquid; likewise they assumed that upon reaching its freezing point only a small subtraction of heat from a liquid was necessary to change the entire mass to a solid. Black realized that these assumptions conflicted with familiar observations: For example, snow masses don't suddenly become raging torrents but melt over an entire summer. Thus he recognized that there are increases in the quantity of heat that are concealed or latent, in the sense of not being registered by a thermometer. Since a mass of ice remains at the melting point for a considerable time before it has all turned to water, the heat it absorbs while at the melting point is latent.

Black quantified latent heat by comparing the time it took to raise a mass of ice from the melting point to 40 °F with the time it took to raise water from 33 °F to 40 °F (1803, 120–122). The former took 21 times as long as the latter; so he calculated the latent heat of ice to be $21 \times (40 - 33) = 147$, from which he subtracted 8 degrees to compensate for the time it took to raise the completely melted ice to its final temperature. His value of 139 °F is only 5° less than the currently recognized value (cf. Holton & Roller, 1958, 332).[15] Black did similar experiments on the latent heat of steam, estimating it to be 810 °F, somewhat less than 972 °F currently recognized (ibid., 334). Black's concept of latent heat permits calculation of heat lost when, for example, a pound of water at, say, 190 °F cools to 0 °F. In cooling to 32 °F a pound of water gives up 1 lb. · 1 °F/lb. · (190 °F – 32 °F) = 158 °F; in freezing it gives up 144 °F, presuming the recognized value; in cooling further to 0 °F, ice gives up 1 lb. · .5 · (32 – 0) = 16 °F, presuming .5 for the heat capacity of ice. The net loss, then, is 318 °F (318 btu = 80.3 kcal). Thus Black provided a simple formula for calculating heat lost and gained during phase transition.

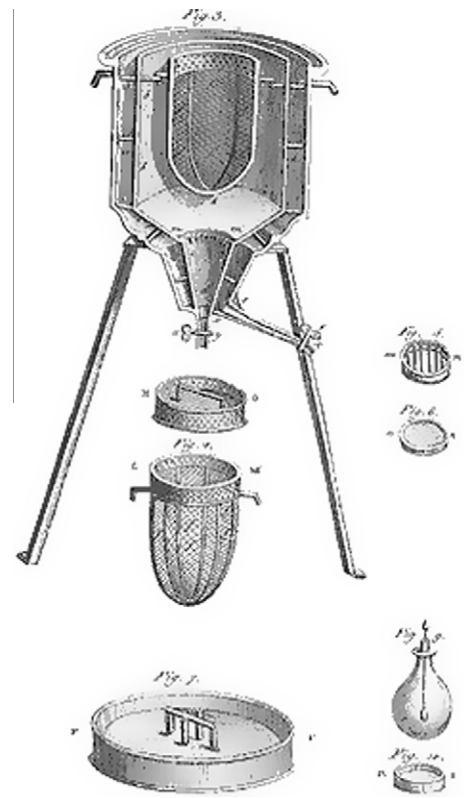

Fig. 5. Ice calorimeter (ca. 1782).

This proved to be the seed of thermochemistry. It made possible Lavoisier and Laplace's ice calorimeter (Fig. 5), though the Frenchmen were not explicit about their use of Black's discovery (Wolf, 1952, I, 184–185). The ice calorimeter measures heat lost in an experiment by weighing the water. Only with Black's concept of latent heat is it possible to transform the water's weight to a measure of heat. Besides providing new methods for determining specific and latent heats, the ice calorimeter made possible measurement of the heat evolved in *chemical* reactions, including combustion and respiration (ibid., 183–188). Latent heat was also important for Watt's study and eventual improvement of the steam engine (McKie & Heathcote, 1935, 34 and 49–50; Wolf, 1952, I, 182).

The quantitative concepts of heat capacity and latent heat did not require Black to determine first that temperature is a continuous quantity. The point of departure for Black's innovations lay in simply *treating* temperature *as* a continuous quantity. His predecessors also treated temperature as a continuous, intensive quantity, and in light of Fahrenheit's improvements, researchers were justifiably *hopeful* that their instruments were capable of measuring temperature. But only Black was able to construct upon his assumption new thermal concepts, useful for explaining and

[15] Like Black's capacity for heat, this too is a relative measure. Black calculated that the quantity of heat it takes to melt a given mass of ice is the same quantity of heat it would take to raise the same mass of water 139 °F. Today, the latent heat of water is given as 144 btu/lb, or 80 kcal/kg.

predicting thermal phenomena. Thereby he established that readings from Fahrenheit's instrument was an instrument whose readings could be subject to arithmetic manipulation and so established the objective validity of a quantitative temperature concept. This is not to claim that in one fell swoop Black solved the practical and theoretical difficulties presented by the temperature concept. Due to the stratification of physical theory, though, the subsequent development of thermodynamics could not meaningfully question whether temperature is a quantitative attribute.

A crucial development was that of Charles (1787) and Gay-Lussac (1802), who found independently that between 0 °C and 100 °C the expansion of gases was proportional to the indications of the mercury thermometer (Holton & Roller, 1958, 369–371). The coefficient of expansion was nearly the same for all gases, and so they proposed that for gases at a constant pressure, volume $\propto$ temperature. Their law had two important consequences. (1) It suggested immediately that temperature could be measured absolutely (i.e., with a non-arbitrary zero). If an increase in volume of $\frac{1}{273}$ accompanies an increase in temperature of 1 °C, then −273 °C constitutes an absolute zero. (2) By combining Boyle's law with the law of Charles and Gay-Lussac, the Ideal Gas law, PV = kT, was derived. That law provided the basis for the thermodynamic temperature scale, today's standard. To be sure, Gay-Lussac expressed reservations about thermometry:

> The thermometer, as it is at present construed, cannot be applied to point out the exact proportion of heat. ... It is indeed generally thought that equal divisions of its scale represent equal tensions of caloric; but this opinion is not founded on any well decided fact. (quoted in Chang, 2004, 57).

But these are reservations about the theoretical explanation of temperature, not reservations about its quantitative status. Indeed, caloric was introduced by Lavoisier in the 1780's as a way to account for Black's discoveries (Holton & Roller, 1958, 334–335), and its plausibility—in spite of the prima facie *implausibility* of an imponderable fluid—lay in the naturalness with which it accounts for the quantitative character of thermal phenomena. As an explanation for thermal phenomena, including the thermometer, caloric—or more generally the idea that heat is a substance—ultimately failed. But the failure of caloric did not count against Black's assumption that temperature is a continuous quantity. Indeed, experiments that refuted the caloric theory took for granted Black's conceptual apparatus. For instance, in order to demonstrate that the heat generated by boring a cannon was not the result of caloric being released, Rumford showed that the specific heat of the chips was the same as the specific heat of the bulk metal (ibid., 338).

The explanatory and predictive power of concepts that *presuppose* temperature's quantitative status reveal the hyperbole in Campbell's claim (1920, 400) that the mercury-in-glass thermometer is as arbitrary as Mohs' scale of hardness.[16] Mohs' scale, unlike Fahrenheit's, plays no role in predicting and explaining mineral phenomena by means of a simple mathematical law. Thus, the *best explanation* for the appearance of the science of heat in the 18th century is that temperature is an intensive, continuous quantity whose levels were, within a modest range, reliably indicated by mercury-in-glass thermometers.

This conclusion suggests a reply to the problem of nomic measurement, as well as a general strategy for showing that a scale is not arbitrary, in the sense of not being merely ordinal. According to Chang the objective validity of the (two fixed-point) mercury thermometer depends on the assumption that mercury expands uniformly with temperature. But, he claims, we can test this assumption only by plotting volume and temperature, which presupposes a reliable thermometer—the very thing in question (59). Justification by means of a direct test is plainly out of the question for the mercury thermometer. But there are less direct ways to justify an assumption. Assuming that mercury expands uniformly with temperature is to assume that temperature, like volume, is a continuous quantity. Generally speaking, if there are higher-level concepts whose power of prediction and explanation presupposes the quantitative character of a lower-level concept then there are good grounds for accepting that the attribute denoted by the lower level concept is quantitative. The power of both specific heat and latent heat, then, are the grounds for maintaining that temperature is a quantitative attribute.

Calorimetry and the discovery of the thermodynamic temperature scale presuppose that temperature is a quantity. In light of the thermodynamic scale it can be shown that, in fact, mercury expands more rapidly as temperature increases. But that discovery does not undercut the work of Black, Lavoisier, Laplace, Charles, and Gay-Lussac. In light of the thermodynamic scale, the expansion of mercury is approximately uniform between 0 °C and 100 °C, certainly more uniform than, say, alcohol or water. If mercury had been scarce or unknown in the 18th century, and Black had had to rely upon alcohol or water as his thermometric substance, it is less likely that the theory of heat would have would have gotten underway. That is, the linear expansion of those substances would have remained unjustified assumptions because neither would have led to higher-level concepts for explaining and predicting thermal phenomena.

In sum, both the claim that temperature is a continuous quantity and the claim that it is measured by Fahrenheit's mercury-in-glass thermometer rest on *abductive* grounds. Their truth, or better, their *approximate* truth, is the best account of Black's success. Black's accomplishments provide a simple and straightforward model for establishing that an attribute is quantitative, a model that compares favorably, I argue, with the models by which measurement theorists propose to determine whether an attribute is quantitative.

5. Representational measurement theory

Michell explains the significance of measurement theoretic models by means of a distinction between the scientific and instrumental tasks of quantification. The scientific task consists in determining *empirically* whether an attribute is quantitative; the instrumental task consists in contriving procedures whereby ratios between levels of an attribute can be reliably estimated, usually by exploiting a relationship between the attribute being quantified and one already quantified (1999, 75). The scientific task

> ... has logical priority in sciences aspiring to be quantitative. In relation to psychology, as far as the logic of quantification is concerned, attempting to complete the scientific task is *the only scientifically defensible* way in which the nexus between the measurability thesis and the quantity objection can be resolved. (76, my italics)

The measurability thesis asserts that some psychological attributes are measurable, while the quantity objection denies that there are any quantitative psychological attributes (25). At least one of these claims is false if, as we may assume, measurable attributes are quantitative. Were they to employ the resources of measurement theory, Michell argues, psychologists could, achieve "a genuine

[16] Barnett refers, similarly, to the "arbitrariness inherent in an elementary temperature scale" (1956, 289). Barnett is referring apparently to the dependence upon a particular thermometric substance and method of graduation of a temperature scale, but there are other senses in which a scale can be arbitrary. E.g., the particular numbers that the scale assigns are arbitrary in the sense that they may be subject to a transformation without affecting the purpose for which the scale is constructed.

resolution of the aporia facing those attempting psychological measurement" (75).

Psychologists have largely ignored measurement theory, and in particular conjoint measurement, which applies to quantities like temperature that don't admit extensive measurement. Conjoint measurement applies to mental attributes *if* they are quantitative; hence, a demonstration that intelligence, say, satisfies (or fails to satisfy) the axioms of conjoint measurement would complete the scientific task of quantification for intelligence. Prima facie, the measurement theoretical method for establishing that a procedure yields a cardinal measure enjoys advantages over the pragmatic method of §4. Judging that a domain of objects satisfies an axiom set is apparently less subjective than judging that treating an attribute as quantitative yields a valuable tool for prediction and explanation. Moreover, by focusing attention on the scientific task, measurement theory enables researchers to avoid hypotheses that suppose, mistakenly, an attribute to be quantitative. But the devil is in the details.

Measurement theory analyzes procedures used to derive *quantitative* measures from *qualitative* observations. In particular, it prescribes the structure that a system of qualitative observations must have in order that numerical relations may be used to *represent*, on the one hand, or *describe*, on the other, those observations in a way that permits mathematical inferences about the empirical situation from which the observations are drawn. Representation differs from description, and the difference reflects a philosophical disagreement over the status of measurement statements. Empiricists hold that measurement statements are abstract, numerical representations of an empirical, *non*-numerical content. Realists, in contrast, hold that measurement statements describe empirical, numerical relations among magnitudes (i.e., levels of an attribute). According to Michell, realists embrace a traditional theory of measurement according to which numbers have "a real-world existence" (2007, 34). Empiricists, he says, regard numbers as "man-made constructions of abstract entities devoid of empirical content" (ibid., 26). I am concerned to show that neither philosophy is satisfactory; for each requires that the empirical world exhibit an *exact* structure, a requirement that flies in the face of our use of empirical predicates. For the sake of concreteness, though, I need to discuss specific axiom sets, and, unfortunately, the different philosophical stances are built into the domains of those axiom sets. In order to avoid the charge that my criticism depends on a particular axiomatization of quantity, §5 scrutinizes systems whose domains are sets of objects or events, while §6 looks at a system whose domain is a set of attributes. In §6 I argue that my criticism can't be avoided by framing the theory in terms of attributes instead of objects or events.

5.1. Weak order

A system of qualitative observations—a so-called 'empirical relational system'—consists of a set of objects (e.g., weights) and a set of observable relations among those objects (e.g., 'x is no heavier than y') (Suppes & Zinnes, 1963, 7). Let the n+1-tuple $\langle D, R_1, \ldots, R_n\rangle$, where D is the domain of objects and R_i the relations, represent an empirical relational system. The *structure* of an empirical system is constituted by the *properties* of its relations (e.g., the relation 'x is no heavier than y' is transitive and strongly connected). The analysis of a measuring procedure is complete once a representation theorem has been proved, i.e., once it has been demonstrated that there is a function that embeds the system of qualitative observations in $\langle N, S_1, \ldots, S_n\rangle$, where N is a set of numbers and S_i numerical relations.[17] By demonstrating an isomorphism[18] between the structure of the observational system and the structure of the numerical system, the theorem guarantees that inferences drawn in accordance with the arithmetic properties of the measures correspond to states of the empirical system.

There are various ways to state he properties of the relations of—i.e., axiomatize—an empirical system capable of ranking the elements of its domain. The variety stems from differences among the relations one chooses to observe. The following axiomatization employs a weak precedence relation **R**, like 'x is no harder than y' (e.g., Krantz et al., 1971, 14).[19]

(i) ∀x∀y∀z (x**R**y & y**R**z → x**R**z)
(ii) ∀x∀y (x**R**y v y**R**x)

R is, in other words, transitive and strongly connected. The corresponding representation theorem demonstrates that any observational system satisfying (i) and (ii) will preserve all the relations in the numerical system $\langle Ra^{+}, \leqslant\rangle$. The only constraint on a function that constitutes an ordinal measurement is that x**R**y iff $f(x) \leqslant f(y)$. An empirical system satisfying (i) and (ii) is a weak order. As long as it can be verified empirically that 'x is no harder than y' is transitive and strongly connected, the system of observations underlying Mohs scale of hardness is a weak order and so suffices for ranking the hardness of a mineral sample.

Measures of hardness are not quantitative because they lack additivity. That is, there are no relations among elements in the observational system that mirror the additive relations in which elements of the numerical system stand to one another. As noted earlier, the fact that the difference between the hardness of diamond and the hardness of quartz equals the difference between the hardness of quartz and the hardness of fluorite can't be said to indicate that diamond is as much harder than quartz as quartz is harder than fluorite. Nor can the average of several measures of hardness be said to indicate a meaningful empirical property.

5.2. Extensive measurement

In order to take advantage of additive relations, an empirical system requires structure beyond that of a weak order. Here too different axiomatizations are possible, depending on the relations one chooses to observe. One familiar system is extensive measurement, which is characteristic of physical science. Weight and length are extensive. Besides a weak ordering relation, extensive measurement requires a concatenation operation, like putting weights in the same pan of a balance or placing measuring rods end to end. Concatenation reduces measurement to *counting* concatenated units, which are equivalent with respect to the ordering relation. However, intensive quantities like temperature, which

[17] Thus, given a measuring procedure (e.g., placing weights in a pan balance and observing whether one side descends) and a system of qualitative observations, <weights; 'x is no heavier than y'>, a representation theorem would show that that this system could be embedded in the system of positive real numbers, $\langle Re^{+}, \leqslant\rangle$. Representation theorems are often accompanied by uniqueness theorems, which show how the functions mapping the observational to the numerical systems relate to one another. For instance, all the functions constituting a temperature scale are positive linear transformations of one another, as in the transformation of Centigrade readings to Fahrenheit readings by $\frac{9}{5} \circ C + 32$. Nothing in my argument turns on the class of mathematical functions to which a scale may or may not belong.

[18] A many-one map (homomorphism) can be traded for a one-one map (isomorphism) by exchanging the domain for equivalence classes of elements of the domain. See Suppes and Zinnes (1963, 26).

[19] 'x is no harder than y' is true iff a sharp point of x fails to scratch a flat surface of y. An equivalent axiomatization is possible by means of a congruence relation, **C** (x is as hard as y), and a precedence relation, **P** (x is less hard than y) (e.g., Hempel, 1952, 59). The congruence relation is transitive, symmetric, and reflexive, while the precedence relation is transitive, **C**-irreflexive (∀x∀y (x**C**y → ¬x**P**y)), and **C**-connected (∀x∀y (¬x**C**y → (x**P**y v y**P**x)). That system can be represented by the numerical system $\langle Ra^{+}, =, <\rangle$.

lack a concatenation operation, are apparently additive too. Thus, concatenation can't be the whole story of physical measurement.

A concatenation operation, $\circ$, is a map from $D \times D$ to D.[20] The following axioms characterize an empirical system capable of extensive measure (e.g., Suppes & Zinnes, 1963, 42).

(i) $\forall x \forall y \forall z$ (x**R**y & y**R**z → x**R**z) (transitivity)
(ii) $\forall x \forall y \forall z$ (x ∘ y) ∘ z**R**x ∘ (y ∘ z) (associativity)
(iii) $\forall x \forall y \forall z$ (x**R**y → x ∘ z**R**z ∘ y) (monotonicity of concatenation)
(iv) $\forall x \forall y$ (¬x**R**y → ∃z (x**R**y ∘ z & y ∘ z**R**x) (solvability)
(v) $\forall x \forall y$¬x ∘ y**R**x (positivity)
(vi) $\forall x \forall y$ (x**R**y → ∃n(n∈Int$^+$ and y**R**nx), where nx is defined recursively as follows: 1x = x and nx = (n − 1)x ∘ x (Archimedean condition)

The representation theorem shows that an observational system satisfying (i)–(vi) can be mapped into the numerical system $\langle Re^+, \leqslant, + \rangle$ (ibid., 43). Besides the constraint for mapping a weak order to a numerical system, a function that constitutes an extensive measurement must also satisfy the equation $f(x \circ y) = f(x) + f(y)$. Measuring weight by means of a pan balance is a paradigm case of an extensive measurement procedure. The domain of the observational system is a set of objects that fit in the pan; the observed relations are 'x is no heavier than y' and 'x and y together are no heavier than z', which corresponds to the concatenation operation of placing x and y in the same pan and observing that the relevant pan does not descend. Measurement theory holds that as long as it can be verified empirically that these two relations satisfy (i)–(vi), the system of observations underlying various weight scales suffice for quantifying weight.

Measurement theory may give a rigorous representation of a measurement procedure, but this alone doesn't justify the axiom set as the touchstone for determining the level (ordinal or cardinal) of that procedure. Consider the domain of an empirical relational system. ''An empirical relational system is a relational system whose domain is a set of identifiable entities, such as weights, persons, attitude statement, or sounds'' is the extent of analysis in Suppes and Zinnes (1963, 7). They might have in mind a set of weights used by a merchant. Provided its elements were well made, such a domain could be shown empirically to satisfy the axioms of a weak order. One could, for example, literally test each ordered triple for transitivity by observing its behavior in a pan balance.[21] But the measurement theoretic analysis of weighing procedures requires more than a set of standard weights; it requires instead an indefinitely large domain of heavy objects. Thus, for Adams, the domain ''consists of the concrete observable things to which numerical measures are to be assigned'' (1979, 208). Apparently, we can only *presume* that the enlarged domain behaves like the merchant's standard weights, because the new domain contains objects, like Mt. Everest, that can't be subject to our measurement procedure. This presumption is not implausible, but let's be clear about its foundation.

The scope of terms like ''identifiable entity'' and ''concrete observable thing'' is vague in comparison with terms like ''positive real number.'' This is not surprising, as we expect borderline cases of empirical concepts while we explicitly rule them out for mathematical ones. Yet a representation theorem requires us to *treat* a collection of concrete observable things *as* a set in the strict mathematical sense. We are required, for instance, to treat all mineral samples as a totality that is clearly enough defined to constitute the domain of a function. Hence, we rule out borderline cases. There is nothing untoward in ruling out borderline cases; doing so is characteristic of applied mathematics. But by ruling out borderline cases we abandon any pretense of arguing by inductive generalization that the objects of the domain satisfy the axioms governing the measurement procedure. This will emerge more clearly as we examine the grounds for claiming that a domain satisfies specific axioms of extensive measurement.

Measurement depends upon observing a specific outcome—a scratch in the case of mineral hardness, a descending pan in the case of weight. Such observations are unproblematic for a trained, well-equipped technician, observing *standard* cases. A sharp calcite point leaves a plainly visible scratch on a flat gypsum surface, and a pan with two equivalent weights plainly descends when the opposite pan contains only a single, similar weight. Outside the standard cases, though, an observer can confront borderline outcomes. Obviously there are perceptual limitations to observing whether a pan has descended. But borderline cases may also be due to an outcome's resembling insufficiently either the paradigm success or the paradigm failure. A successful scratch of kyanite, for example, depends upon the direction in which the scratch is attempted.[22] Although borderline cases are typical of empirical concepts, they need not undermine theories that employ mathematical inference, as one can always adopt conventions to deal with borderline cases (e.g., the pan does not descend unless it plainly descends). But conventions take us beyond observation. This is nicely exhibited in the decision to treat 'x is no heavier than y' as transitive in spite of counterexamples: Without this idealization, we forfeit the convenience of arithmetic. It is true that as measurement procedures become more precise, the counterexamples to transitivity diminish. But this shouldn't be taken to mean that convenience could disappear in principle from the task of verifying that the axioms of extensive measurement have been satisfied. Two of extensive measurement's axioms remain problematic even if we treat the axioms as describing, in Michell's phrase, ''the form the data *would have . . . were they completely free of error*'' (Michell, 2007, 31). The sort of idealization demanded by both solvability and the Archimedean condition is not a matter of error free data.

Solvability postulates, for example, that elements of the domain are always available to make up the difference between a heavier object x and a lighter y; that is, an element z can always be found such that its concatenation with y *exactly* balances x. The pragmatic motivation for such a principle is clear: By postulating the object z, we lay claim to the empirical correlate of a number equal to the difference between x and y. Given the amorphous character of empirical domains, and the ease with which borderline cases can be handled, this postulate is not hard to swallow. But our confidence in solvability is not based on having successfully located objects that exactly make up the difference between a pair of objects. Rather, our confidence is based on the benefits derived from treating an empirically given domain as though relations among its objects imitate the relations among objects that constitute mathematical continua. Likewise, the same benefits encourage us in our conclusions about the world of error free data. Abductive reasoning is at work here. The *best explanation* for the benefits of treating a domain as though it satisfies the axioms of extensive measurement is that objects in that domain satisfy those axioms, at least approximately. That is, our justification for applying real numbers to the domain is abductive.

The Archimedean condition is likewise impossible to verify by simply observing the behavior of the domain. Given any element x of the domain, the condition postulates as many exact copies of

[20] Introducing an operation need not violate the stipulation that an empirical system consists of a domain and a set of relations; for a binary operation is a special case of a ternary relation by identifying ∘ with the ternary relation x ∘ y = z, provided x ∘ y = z & x ∘ y = w → z = w (Suppes & Zinnes, 1963, 5).

[21] Strong connectedness would hold by definition.

[22] Scratched in one direction, kyanite falls between 4.5 and 5 on Mohs scale; in the perpendicular direction, though, it falls between 6.5 and 7.

x as one needs to produce, via concatenation, an object which exceeds any other element. This axiom is also easy to swallow, as an extensive measurement procedure depends upon an unlimited supply of identical units. But no one would even begin to check by direct observation that every object in the domain comes equipped with indefinitely many exact copies. Here too the anticipation of useful mathematical manipulation—rather than empirical facts about the domain—informs the conditions of extensive measurement. Like solvability, the Archimedean condition idealizes by creating objects that eliminate inexactness from the empirical world. Here too, our belief that a particular domain approximately satisfies this axiom does not rest upon an inductive generalization from observations of the domain. There is no question, for instance, of sampling randomly from the domain of the empirical relational system. Here too, the warrant is abductive: The best explanation for the success of the measurement procedure is that the objects of its domain stand, more or less, in the relation specified by the axiom.

5.3. Conjoint measurement

Conjoint measurement is an advance in measurement theory that occurred a half century ago (Luce & Tukey, 1964). The axioms of conjoint measurement gave, for the first time, conditions under which interval and ratio scales can be constructed from strictly ordinal data (Krantz et al., 1971, 245ff.). Conjoint measurement gives, then, conditions under which a thermoscope can be used as a thermometer, and in this way it provides a rigorous mathematical solution to the problem of nomic measurement. Since psychologists, too, must measure attributes for which no obvious concatenation operation exists, Michell's plan to found psychological measurement upon conjoint measurement has obvious merit.

In its simplest form, conjoint measurement applies to situations involving three variables, one of which, P, is a function of the other two, A and X; i.e., P = f(A, X). The relationship between mass, on the one hand, and density and volume, on the other, is one such situation (Michell, 1990, 69); another is Michell's hypothetical case in which performance is a function of ability and motivation (1999, 201ff.). The restrictions on P, A, and X are minimal. P must be an ordinal variable with infinitely many values, but A and X may be merely classificatory. An empirical relational system that satisfies these conditions is conjoint. Not all conjoint systems are quantitative, like the mass/density/volume system. However, for a conjoint system in which $\succsim$, the ordering relation on P, satisfies (i) double cancellation, (ii) $\tilde{\tilde{s}}$olvability, and (iii) the Archimedean condition, P, A, and X are *all* quantitative variables and f is a non-interactive function, i.e., A and X affect P independently of one another. This is the substance of the representation theorem for an additive conjoint measurement (Krantz et al., 1971, 257ff.).

Conjoint measurement, especially the double cancellation axiom, is rather less intuitive than extensive measurement. Fortunately, for our purpose it suffices to note the ideal character of conjoint measurement and the occurrence of both solvability and the Archimedean condition.[23] Plainly a conjoint system depends upon idealization in exactly the ways an extensive measurement does. Here too, our belief that a particular domain approximately satisfies a conjoint system does not rest upon an inductive generalization from observations of the domain. It rests, instead, upon an objective inference from the benefits of treating the objects as though they satisfy the axioms.

Measurement theorists are not oblivious to the gap between domains that confront empirical scientists and the domains of empirical relational systems. Krantz et al. write, "The axioms purport to describe relations, *perhaps idealized in some fashion*, among certain potential observations ..." (1971, 26, my italics). But they haven't, I believe, stressed sufficiently that the idealizations arise from emulating the very numerical system whose structure is to be proved isomorphic to that empirical relational system. Thus, a representation theorem establishes at most the existence of a function from a *quasi-empirical* relational system to a numerical relational system. Here a quasi-empirical system is a set-theoretic counterpart of a domain of empirical investigation, constructed by substituting exact mathematical domains and concepts for inexact empirical ones. There can be no structural isomorphism between a *genuinely empirical* relational system—i.e., a system whose concepts are strictly empirical—and a numerical relational system. For empirical concepts, which are inexact, obey a different logic from mathematical concepts, which are exact. Because they admit borderline cases, for instance, empirical concepts don't obey the law of excluded middle.[24] In truth, a representation theorem describes the idealized assumptions we bring to a genuine empirical relational system in order to produce a hypothetico-deductive system that can take advantage of the inferential powers of arithmetic. Making explicit our idealizations is a worthwhile project, but it's misguided to characterize an empirical relational system, as Michell does, as presenting "observable, surface structures enabling tests of features of quantitative structures" (2007, 36). Surface structures are quite the opposite of idealized structures.

Instead of *sharing* the structure of a numerical relational system, a genuine empirical relational system is *identified* with a quasi-empirical system and the latter shares the structure of a numerical system. Identification consists in replacing inexact concepts by exact ones. For instance,

> [A] basic statement of an equality involves the replacement of an empirical non-transitive relation by the mathematical relation of equality. (Körner, 1964, 283)

But replacement is not always as simple as replacing a concept that admits borderline cases with a concept that has sharp edges. Thus, exchanging empirical continuity for a concept suitable to mathematical inference takes more than replacing each inexact element of the concept with an exact counterpart.

> A basic statement of continuous transition involves the replacement of an empirical notion of continuity, definable in terms of a finite set of inexact classes, by a mathematical notion of continuity, defined in terms of a non-denumerably infinite set of exact classes. (ibid., 282–283)[25]

Identification, then, is a matter of choosing to treat a thing of one sort as a thing of another sort in order to achieve, in Körner's phrase, deductive unification. The warrant for identifying a domain of empirical investigation with a quasi-empirical domain is, obviously, the pragmatic value in doing so. Since ancient times, physical science has treated empirical predicates as exact, quasi-empirical

[23] Solvability for a conjoint system is uncomplicated: Given three of four values a_i, a_j, x_i (or a_i, x_i, x_j) *there exists* a fourth, x_j (a_j) such that $a_i x_i = a_j x_j$. I.e., a solution to the preceding equation can always be found. Solvability is not directly testable (Michell, 1990, 79). The case is similar for the Archimedean condition, which is much more complicated to state; it too makes an existential claim whose justification faces limitations in time and technology.

[24] Körner (1964) shows in detail that empirical predicates obey a modified two-valued logic different from the classical two-valued logic.

[25] See Körner (1962) for a full discussion of the differences between an empirically continuous series and a mathematical continuous one. An empirically continuous series, i.e., one tied to direct observations, is not dense in the mathematical sense, and so lacks additive structure; however, it is empirically dense, a notion defined in terms of borderline cases.

ones, though with hardly a thought. This maneuver is already implicit in pedestrian applications of mathematics. Numerals are applied, in the first place, as inexact, empirical predicates, the results of counting or measurement.[26] Since empirical predicates like "two" and "one and a half meters" admit borderline cases, they must be identified with exact, mathematical ones in order to take advantage of the inferential powers of arithmetic. Thus, some measurement statements are descriptive and others are merely representations. The former involve inexact, empirical predicates, while the latter involve exact, non-empirical predicates. Only the latter are employed in prediction and explanation.

The quasi-empirical character of empirical relational systems undermines the idea that measurement theory provides observable surface structures capable of refuting the hypothesis that a given procedure yields a cardinal measurement. For predictive and explanatory success will generally trump observations purporting to show that an attribute is not quantitative. One could, I suppose, appeal to measurement theory to explain why a particular procedure failed to yield quantitatively valuable results. But even here the failure of the domain to satisfy the axioms could be due to technological limitations, and someone with faith in the procedure would be unmoved. Likewise, the quasi-empirical character of measurement theory puts the lie to the proposal that observable surface structures could, in the absence of predictive and explanatory success, establish the level of a measurement procedure. The level of a measuring procedure rests, then, on a pragmatic foundation, not an observational one. New procedures should be judged pragmatically, and the point of our historical investigation of the temperature concept is to provide a concrete paradigm for such judgments.

6. Realist measurement theory

So far, the critique of measurement theory as the touchstone of quantity depends upon an empiricist or representational formulation of the theory. In order to argue that the example of thermometry is a more suitable touchstone, it's necessary to show that an alternative formulation of measurement theory, one that escapes problems arising from a limited domain of objects, is equally problematic.

Michell argues for a realist version of measurement theory that contrasts with the representational version (Michell, 2005, 2007). The realist wants to avoid the idealizations that infect representational measurement theory with 'a slavish imitation of the number system' (Michell, 2007, 33–34; cf. Suppes & Zinnes, 1963, 45). In the same vein he hopes to avoid treating numbers as abstract entities, i.e., as entities existing outside space and time. Michell objects, in particular, to employing entities (i.e., numbers) that are "related externally to features of [an empirical] situation by human convention" (2005, 287). The point of realist measurement theory, then, is to explain measurement in a naturalistic fashion such that "the realm of space and time is world enough" for understanding nature (Michell, 2007, 33).

Representationalists are not anti-realist, and indeed empiricism of the last 60 years is staunchly realist. They believe, like Michell, that the truth of statements of science and mathematics requires objects that are *independent* of observation and correctly *described* by those statements (Michell, 2004, 286). Thus abstract objects, in particular, sets, were established in the philosophical landscape thanks to the indispensability argument of Quine and Putnam (Quine, 1961, 1–19 and especially Putnam, 1971, 337ff.). According to that argument, our ontology requires mathematical objects because they are indispensable for science.

> If the numericalization of physical magnitudes is to make sense, we must accept such notions as function and real number. . . . Yet if nothing really answers to them, then what at all does the law of gravitation assert? For that law makes no sense at all unless we can explain variables ranging over arbitrary distances (and also forces and masses, of course). (Putnam, 1971, 341)

Putnam turns to abstract objects, because it is preferable to having to postulate the existence of an actual infinity of physical objects (339). Empiricists are not entirely comfortable about abstract objects, because it is problematic to say how humans come to know about them (Benacerraf, 1973). Quine, whose views on metaphysics and epistemology have been dominant since the collapse of logical positivism, solves the problem by undermining the special status of physical objects. He maintains that one's ontology consists entirely of entities—physical objects, irrational numbers, etc.—postulated in order to "round out and simplify our account of the flux of experience" (1961, 18). We are committed to just those entities that do the best job of rounding out and simplifying. Thus, Quine points out

> . . . no measurement could be too accurate to be accommodated by a rational number, but we admit the [irrationals] to simplify our computations and generalizations. (1986, 400)

Empiricists accept the existence of mathematical objects, then, on pragmatic grounds similar to the grounds presented in §4 for the quantitative temperature concept.

Michell finds a 'worldly' alternative to representational measurement theory in Hölder's axioms for an unbounded continuous quantity, the domain of which is all possible levels of an attribute. Such domains are "better suited, conceptually, to account for measurement . . . if only because no actual set of rigid, straight rods could ever instantiate *all possible lengths*" (33, my italics). Michell formulates Hölder's axioms for an unbounded continuous quantity Q, with levels a, b, c, d, as follows.

1. Given any magnitudes, a and b, of Q, one and only one of the following is true:
   (i) a is identical to b (i.e., $a = b$ and $b = a$);
   (ii) a is greater than b and b is less than a (i.e., $a > b$ and b; or
   (iii) b is greater than a and a is less than b.
2. For every magnitude, a, of Q, there exists a b in Q such that $b < a$.
3. For every pair of magnitudes, a and b, in Q, there exists a magnitude, c, in Q such that $a + b = c$.
4. For every pair of magnitudes, a and b, in Q, $a + b > a$ and $a + b > b$.
5. For every pair of magnitudes, a and b, in Q, if $a < b$, then there exists magnitudes, c and d, in Q such that $a + c = b$ and $d + a = b$.
6. For every triple of magnitudes, a, b, and c, in Q $(a + b) + c = a + (b + c)$.
7. For every pair of classes, ϕ and ψ, of magnitudes of Q, such that
   (i) each magnitude of Q belongs to one and only one of ϕ and ψ;
   (ii) neither ϕ nor ψ is empty; and
   (iii) every magnitude in ϕ is less than each magnitude in ψ, there exists a magnitude x in Q such that for every other magnitude, x', in Q, if $x' < x$, then $x' \in \phi$ and if $x' > x$, then $x' \in \psi$ (x may belong to ϕ or ψ). (Michell, 1999, 51–53; cf. 2007, 20–21).[27]

[26] The criteria for applying inexact numerical predicates can presumably be given in terms of inexact, qualitative predicates. But owing to their inexactness those criteria would not constitute an axiomatization.

[27] In the 2007 formulation Michell restricts Q to quantitative attributes for which "$a + b = c$ iff c is entirely composed of discrete parts a and b."

Hölder's axioms, says Michell, specify the structure that the levels of an attribute "must possess if any two of them are to stand in ratios which can be expressed numerically" (1999, 59). By explaining[28] how different relational systems (extensive, conjoint, etc.) sustain a ratio scale they serve as the deep theoretical structure underlying measurement (208). And thus Michell relegates empirical systems to presenting "the kinds of observable, surface structures enabling tests of features of quantitative structure."

The realism of Michell's alternative lies in the commitment to the existence of all possible levels of an attribute, viz., the entities that Hölder's axioms seek to describe. If, as Michell maintains, the levels of an attribute (and relations between them) exist in space and time, then numbers, which are ratios between magnitudes, are part of the natural world. In that case there is no need to invoke abstract entities (1999, 59–62).

Some will be skeptical of invoking systems of attributes to deal with there being "no actual set of rigid, straight rods [to] instantiate *all possible lengths*." In the first place, if there are unactualized possible lengths, then there are presumably objects outside space and time. According to Hölder's scheme, length is an unbounded, continuous quantity, and so there is no greatest length. But if contemporary cosmology is right, the universe is finite, and so there must be levels of length that are not instantiated in space and time. Michell seems to face the same problem as the representationalist—too few actual entities to fill the theory's domain. Secondly, attributes have been a source of philosophical discomfort since the days of Plato and Aristotle. While Plato 'located' attributes (i.e., forms) outside space and time, Aristotle scoffed at the suggestion, insisting forms must be immanent. Michell follows Aristotle's worldly option.

> While any specific attribute can only ever exist at some particular spatiotemporal location as a feature of something or other, it owes no logical allegiance to any location and in this sense is *general*. For this reason attributes are called 'universals'. (2005, 286)

Plato and Aristotle agreed on both the reality and the generality of attributes. While Plato appealed to transcendence to explain how distinct objects could share the same attribute, Aristotle claimed that the *same* form (humanity, say) could be an ingredient in *distinct* individuals. Both explanations are implausible. Michell hopes to avoid their troubles by making attributes particulars whose generality arises from 'owing no logical allegiance' to particular spatiotemporal locations. However, if Plato's humanity (weight, temperature, etc.) owes no logical allegiance to Plato or his spatiotemporal location, it could have been a feature of Aristotle or his spatiotemporal location. But how does one individuate Plato's humanity from Aristotle's, if not by reference to the concrete particulars in which they are instantiated? The difficulty of individuating attributes and possibilia prompts Quine to reject them, of course (1961, 4 and 10). Furthermore, Quine's ontology of physical objects and sets renders attributes unnecessary for scientific discourse.

But Michell finds additional advantages in a realist measurement theory. He notes that although "axiom systems, like those for extensive and conjoint systems, obtain for all attributes measured in physics, ... no one in the history of physics may ever have made observations relating directly to the axiom systems for extensive or conjoint measurement" (2007, 32). Representational measurement theory cannot explain this anomaly, but realist measurement theory can.

> Take the case of length. If we come to the above example of an empirical extensive system armed with a concept of length as an unbounded continuous quantity, in the spirit of Hölder's axioms, then, knowing what we do about rigid straight rods and their behavior in standard circumstances, we have no trouble inferring that the six axioms given above[29] are true. (32–33)

That is, Hölder's axioms explain why physical scientists feel no urgency to test whether the properties they investigate satisfy the axioms of an empirical relational system: An axiomatization of length in the spirit of Hölder—and our experience with rigid rods—imply that the axioms of extensive measurement hold for rigid rods. Our acceptance of Hölder's theory of continuous quantity also explains our inclination to blame the data rather than the axioms when observations appear to falsify the axioms for an empirical relational system (33). Michell hastens to note, "Of course, in accord with the values of science, this theory is ultimately based upon observational evidence, but it is not based upon direct tests of these six axioms or of any others" (ibid.). This claim is crucial to Michell's preference for realist over empiricist formulations of measurement theory. For it implies that only the former claim the authority of direct empirical tests.

Empiricists don't attach the same importance as Michell to direct empirical tests. They insist on evaluating observational evidence holistically, and so consider as well the practical value of the hypothesis being tested. An empiricist could embrace a concept of length in the spirit of Hölder's axioms, but not on observational grounds alone. In the empiricist tradition of the last 60 years, then, the domain of Hölder's axioms are in the same position as the domain for the axioms for extensive measurement: Their domains, once they are postulated, stand or fall with respect to the benefits they offer toward understanding nature.

Michell is suspicious of postulation, and he reminds readers that postulation shares all the advantages of theft over honest toil (2007, 24). But I don't see that a realist program has any choice but to postulate attributes (as well as sets) and argue for the pragmatic value of doing so. This is clearest in view of the theoretical role of Hölder's axioms. They are significant because they permit a mathematical proof that an unbounded, continuous quantity shares the structure of the positive reals. The proof requires that a finely articulated structure be packed into axioms. It is no accident, for example, that the levels of a continuous attribute mirror both the density and the completeness of the reals. Indeed, Hölder's seventh axiom is identical to Dedekind's axiom of completeness, except that sets of attributes are substituted for sets of rationals. Structure like that cannot be put to direct empirical test because our observations are not that fine-grained. Thus, empirical facts about the domain can confirm Hölder's axioms only in the light of the pragmatic value of the mathematics whose structure Hölder's axioms imitate. The slavish imitation that concerns Suppes and Zinnes is equally a problem for Michell.

The realist formulation of measurement theory is no more closely tied to observation than its empiricist rival, and both versions are guilty of tailoring their axioms for the sake of representation theorems. Hölder's axioms, like those of Suppes and Zinnes, involve exact concepts; both involve quasi-empirical structures. Thus, empirical tests of either involve identifying inexact concepts with exact ones. And since we do not observe instances of exact concepts, the grounds of identification are, therefore, pragmatic.

Michell will reject my distinction between empirical and quasi-empirical for ignoring the transcendental ground (in Kant's sense) of measurement. According to Michell, "The fact that that measurement of physical attributes occurs at all implies that there must be something about the character of the relevant attributes that makes it possible" (2007, 33). Measurement rests "upon the possibility of real-world systems being *similar in structure* to mathematical systems" (34, my italics). As Hölder's theorem demonstrates that the

[28] The explanations substitute the (weaker) Archimedean condition for Holder's 7th axiom (Michell, 1999, 211).

[29] The reference is to the Suppes and Zinnes axioms for extensive measurement.

real number structure is shared by the ratios of magnitudes of an unbounded continuous quantity, Hölder's axioms, far from being quasi-empirical, describe what the empirical world *must be like* in order for measurement to be possible. Undoubtedly structural similarity is necessary for mathematics to be applicable to nature; but structures can be similar without being isomorphic. Provided length, weight, etc. behave *approximately* as Hölder's axioms prescribe, then the mathematics of continua will be applicable.[30] Structural similarity in the sense of mathematical isomorphism is too strong a requirement, and the situation in contemporary physics supports this claim.

A fundamental problem in contemporary physics is unifying quantum mechanics with general relativity. Quantum field theories have successfully accounted for three of the four fundamental forces, electromagnetic, and weak and strong nuclear forces. But, as Maddy puts it, gravitational force ''has an annoying habit of generating impossible (infinite) values'' (1992, 285). Physicists propose various strategies for dealing with these anomalies, among them abandoning the continuity of space and time. Thus, Feynman writes,

> I believe that the theory that space is continuous is wrong, because we get these infinities and other difficulties . . . (quoted in Maddy, 1992, 285)

And another physicist, Isham, speculates,

> . . . it is clear that quantum gravity, with its natural Planck length, raises the possibility that the continuum nature of spacetime may not hold below this length, and that a quite different model is needed. (ibid.)

These physicists consider it a genuine possibility that length and distance are not continuous quantities. But presumably neither would claim that the correctness of their speculations would reveal that time-tested methods for measuring length and distance were a sham. For the purpose of understanding quantum gravity, a model in which length and distance are continuous quantities may not be appropriate. But the shortcomings of the continuous model wouldn't cast doubt upon previously successful methods of measuring length and distance. Success suffices to defend these methods, and that success is explicable as long as length and distance are *approximately* continuous.

Models enable us to draw inferences in accordance with mathematics, i.e., they achieve deductive unification. This they accomplish by means of exact concepts. Black advanced the science of heat by modeling temperature as a continuous quantity, i.e., by identifying readings from the thermometer with a particular set of exact concepts. His advances were possible because temperature approximates a continuous quantity, and it approximates a continuous quantity insofar as there are tangible benefits to treating it as continuous. In sum, the structural similarity that makes applied mathematics possible is a weaker notion than the one Michell needs in arguing that the world must conform to Hölder's axioms if measurement is to be possible. Satisfaction of something like the conditions of empirical continuity in Körner (1962) will suffice for measurement.

7. Conclusion

Michell dismisses the pragmatic defense of treating psychological attributes as quantitative (1999, 20ff; cf. 217ff.). When Lord and Novick argue that, ''To the extent that this scaling produces a good empirical predictor the stipulated interval scaling is justified'' (Lord & Novick, 1968, 22), he responds that they are ignoring the fundamental scientific issue of whether or not an hypothesized attribute is quantitative.

> Only when such a theory has been subjected to some experimental test sensitive to the presence or absence of quantitative structure in the hypothesized attribute can any conclusions be drawn about whether or not test scores are interval scale measures of anything. Weaker tests, such as the test scores being a good predictor of related criteria, are not sensitive to the presence or absence of quantitative structure in the underlying attribute because no matter which way they turn out they cannot rule out the hypothesis that this attribute is quantitative. (1999, 21)

When Michell observes that predictive failure of weaker tests won't rule out the hypothesis that an attribute is quantitative, he implies that the stronger tests made possible by measurement theory could rule out the hypothesis that an attribute is quantitative. But it's a myth to think such a test might once and for all rule out an hypothesis of quantity. On the one hand, failure to satisfy the axioms for quantity could be blamed upon the testing instrument, as would have been the case had alcohol thermometers been used instead of mercury thermometers. On the other hand, the exact, quasi-empirical character of the axioms, in contrast to the inexact, empirical character of observations, gives psychologists an avenue on which to shunt contrary data. If treating an underlying attribute, say intelligence, as quantitative allows one to employ mathematics to bring order to the data, then a plausible explanation for this success is that the attribute is approximately quantitative. If the same presumption leads to higher-level concepts, analogous to Black's specific and latent heats, then it's hard to imagine what additional scientific purpose could be served by bringing Hölder's axioms to bear.

Acceptance of the pragmatic approach to quantification does not, of course, entail the existence of quantitative psychological attributes. The pragmatic value must be demonstrated. The conceptual advance that we admired in temperature measurement has no obvious analogue in, for instance, intelligence measurement. The general intelligence factor, g, does not occur in psychological laws in the straightforward way that temperature occurs in $m_1c_1(t_1 - t_f) = m_2c_2(t_f - t_2)$ and $\Delta Q = mc\Delta t$. g correlates mathematically with academic success, income, etc., but it is not clear that the correlation provides greater understanding when g is treated as cardinal rather than ordinal data. This issue could be settled affirmatively by calling attention to a conceptual advance analogous to Black's introduction of specific and latent heats. The statistical character of psychological concepts and laws will, however, complicate any attempt to demonstrate such an analogy. That project is better saved for another occasion.

Michell blames psychologists' acceptance of the measurability thesis upon their unreflective commitment to measurement as the assignment of numerals to objects or events according to rules (1999, 20–21; cf. Stevens, 1951, 1). By failing to exclude any attributes from the practice of measurement this definition of measurement cancels the scientific task (1999, 77; 216).[31] The history of temperature suggests a middle path between embracing a research

[30] According to Michell, well-known questions about 'the unreasonable effectiveness of mathematics' (cf. Wigner, 1960) result from representational accounts of measurement (2005, 292). But neither Wigner nor anyone else is puzzled about the effectiveness of real numbers, which are idealizations of empirical observations. Wigner's examples of unreasonable effectiveness are, rather, characterized by our getting something out of the mathematics that we did not put in (10). Thus complex numbers, which are crucial for quantum mechanics, were not suggested by physical observations but chosen ''for their amenability to clever manipulations'' (7). In fact, none of Wigner's examples involve measurement.

[31] There is some basis for this charge but it is misleading. While Stevens accepts coding (e.g., 0 = female, 1 = male) as measurement, he distinguishes different types of measurement by the types of empirical operations necessary to assign a numeral to ''the aspects of objects'' (1951, 23). Stevens agrees with Michell that ''quantity'' is appropriate *only* for interval and ratio scales (27).

program dictated by measurement theory and simply abandoning the scientific task of quantification. It is open to psychologists to be encouraged *or discouraged* in their attempts to quantify mental attributes by comparing their accomplishments with Black's.[32]

References

Adams, E. (1979). Measurement theory. *Current research in philosophy of science*. Philosophy of Science Association (pp. 207–227).

Aristotle (1941). In McKeon (Ed.), *The basic works of Aristotle*. New York: Random House.

Barnett, M. (1956). The development of thermometry and the temperature concept. *Osiris, 12*, 269–341.

Benacerraf, P. (1973). Mathematical truth. *Journal of Philosophy, 70*, 661–679.

Black, J. (1803). Expansion-thermometers. In J. Robison (Ed.), *Lectures on the Elements of Chemistry*. London: Longman and Rees.

Campbell, N. (1920). *Physics, the elements*. Cambridge: Cambridge University.

Chang, H. (1995). Circularity and reliability in measurement. *Perspectives on Science, 3*, 153–172.

Chang, H. (2004). *Inventing temperature*. Oxford: Oxford University.

Clagett, M. (1959). *The science of mechanics in the middle ages*. Madison: University of Wisconsin Press.

Crosby, A. (1997). *The measure of reality*. Cambridge: Cambridge University Press.

De Luc, J. (1772). *Recherches sur les modifications de l'atmosphère* (Geneva: n.p.).

Dulong, P. and Petit, A. (1817). Recherches sur la mesure des tempèratures et sur les lois de la communication de la chaleur. *Annales de chimie et de physique, 2d ser.*, 7, 113–154, 225–264, 337–367.

Galilei, G. (1974). *Two new sciences* (trans. Drake). Madison: University of Wisconsin Press.

Hempel, C. (1952). *Fundamentals of concept formation in empirical science*. Chicago: University of Chicago Press.

Hölder, O. (1901). *Die Axiome der Quantität und die Lehre vom Mass* (translated in Michell and Ernst (1996–1997)).

Holton, G., & Roller, D. (1958). *Foundations of modern physical science*. Reading, MA: Addison-Wesley.

Klein, H. (1974). *The science of measurement*. New York: Simon and Schuster.

Körner, S. (1962). On empirical continuity. *The Monist, 47*, 1–19.

Körner, S. (1964). Deductive unification and idealisation. *British Journal for the Philosophy of Science, 14*, 274–284.

Krantz, E. et al. (1971). *Foundations of measurement* (Vol. I). New York: Academic Press.

Kuhn, T. (1961). The function of measurement in modern physical science. In Woolf (Ed.), *Quantification* (pp. 31–63). Indianapolis: Bobbs-Merrill.

Lord, F. (1953). On the statistical treatment of football numbers. *The American Psychologist, 8*, 750–751.

Lord, F., & Novick, M. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison Wesley Publishing.

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

Maddy, P. (1992). Indispensability and Practice. *Journal of Philosophy, 89*, 275–289.

Maier, A. (1982). *On the threshold of exact science* (ed. and trans. Sargent). Philadelphia: University of Pennsylvania.

McGee, T. (1988). *Principles and methods of temperature measurement*. New York: John Wiley and Sons.

McKie, D., & Heathcote, N. (1935). *The discovery of specific and latent heats*. London: Edward Arnold and Co.

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

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

Michell, J. (2004). The logic of measurement: A realist overview. *Measurement, 38*, 285–294.

Michell, J. (2007). Representational theory of measurement. In Boumans (Ed.), *Measurement in economics: A handbook* (pp. 19–39). Amsterdam: Academic Press.

Michell, J., & Ernst (1996–1997). The axioms of quantity and the theory of measurement, Parts I and II. *Journal of Mathematical Psychology, 40*, 235–252 and *41*, 345–356.

Middleton, W. (1966). *A history of the thermometer and its use in meteorology*. Baltimore: Johns Hopkins.

Murdoch, J. (1974). Philosophy and the enterprise of science in the later middle ages. In Elkana (Ed.), *The interaction between science and philosophy*. Atlantic Highlands, NJ: Humanities Press.

Oresme, N. (1968). *Tractatus de configurationibus qualitatum et motuum* (trans. Clagett). Madison: University of Wisconsin Press.

Putnam, H. (1971). Philosophy of logic. *Mathematics, matter, and method*. Cambridge: Cambridge University Press (pp. 323–357).

Quine, W. (1961). *From a logical point of view* (2nd ed., rev.). New York: Harper and Row.

Quine, W. (1986). Reply to Charles Parsons. In Hahn & Schilpp (Eds.), *The philosophy of W.V. Quine* (pp. 396–403). LaSalle, IL: Open Court.

Roche, J. (1998). *The mathematics of measurement*. London: The Athlone Press.

Roller, D. (1966). The early development of the concepts of heat and temperature. *Harvard case histories in experimental science*. Cambridge, MA: Harvard University Press (pp. 119–213).

Schooley, J. (1986). *Thermometry*. Boca Raton: CRC Press.

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

Suppes, P., & Zinnes, J. (1963). Basic measurement theory. In Luce et al. (Eds.), *Handbook of mathematical psychology* (pp. 1–76). New York: John Wiley and Sons.

Taylor, F. (1942). The origin of the thermometer. *Annals of Science, 5*, 129–156.

Toulmin, S. (1960). *The philosophy of science: An introduction*. New York: Harper and Row.

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

Wolf, A. (1950) (2nd ed.). *A history of science, technology and philosophy in the 16th and 17th centuries* (Vol. 1). New York: Macmillan Company (First edition printed in 1935).

Wolf, A. (1952) (2nd ed.). *A history of science, technology and philosophy in the 18th century* (Vol. 1). New York: Macmillan Company (First edition printed in 1938).

[32] This paper benefitted from comments by two anonymous referees and conversations with Joel Michell, Mpho Selomo and Matt Fahy.