Procedures for the Precise Measurement of Online Cognitive and Cultural Processes[1]

Joseph Woelfel

University at Buffalo

State University of New York

Copyright
May 23, 2010

"...science is the observation of phenomena and the communication of the results to others, who must check them" --Niels Bohr.

A measurement is a comparison to a standard. -- William Shockley

Science is a way of trying not to fool ourselves.   -- R. P. Feynman

---

The Problem

        Science does not progress steadily forward, but makes mistakes, branches, retreats and starts again, but, through relentless observation, communication and checking, corrects its errors and moves on. Since the advent of modern science, we have passed through numerous serious errors, many of which have been discovered and left behind. Only a few of the most notorious errors are the luminiferous aether, phlogiston, Lysenkoism, Lamarckism, the geocentric universe, Ptolemy's solar model, the brain as radiator, nerves as hollow tubes conducting spirit, stars as fires, Atlas on a turtle holding up the world, the sun as a huge fire on a chariot circling the earth, the indivisibility of atoms, the idea that objects seek their proper place and have their own proper motion, the spontaneous generation of flies, the perfection of the heavenly spheres, the corpuscular theory of light, the flat earth hypothesis, the elemental particle theory (i.e., that all things are made from earth, air, fire, water), and Aristotle's dynamic theory of motion, among many, many more.

        Social science, on the other hand, has generated a plethora of overlapping and contradictory theories of human thought and behavior, such as Aristotle's rational actor model, the free will model, the Calvinist predestination model, the needs and gratifications approach, the Freudian psychodynamic model, Jung's theory of cultural archetypes,[2] cognitive dissonance theory, the Wisconsin status attainment model (Haller 2000), Marxism, Capitalism, Weber's Ideal Type model, the Galileo model, symbolic interaction theory, Lewin's field theory, Pareto's circulation of the elite, Durkheim's theory of the collective consciousness, Parson's AGIL model, the damped harmonic oscillator model, stimulus-response theory, contiguity of response theory, labeling theory and many, many more. Since the advent of modern social science, the number of these competing theories that have been decisively rejected is -- none.

        Why is it that the physical sciences have been so successful at culling unsuccessful theories, while the social sciences have failed to eliminate *any* of their competing theories? Are the social sciences, as Marvin Gardner once said of psychology, a fast race around a short, round track?

        The conventional answer, of course, is that the social sciences are much more difficult than the physical sciences due to the nature of the subject matter. Physical phenomena are concrete, specific and easily observable; social and cognitive phenomena are inherently vague and evanescent, spontaneous and not governed by natural "laws." Scientists of any kind, however -- particularly those who understand that the result of observation is not independent of the way observations are carried out -- ought to be deeply suspicious of words like "inherent".

        The notion that the subject matter of social science is inherently different from that of physical science, that we know this independent of any observations or measurements, and that

_____

[2] Chrysler helped pave the way to bankruptcy by relying on Carl Jung's theory of cultural archetypes to design and sell vehicles in the 1990's.

such differences require radically different concepts of measurement is an extraordinary claim, and, as such, ought to require extraordinary evidence. Yet the serious scholar will look in vain for any evidence whatever, for this notion is not a scientific finding, but a philosophical belief. Like Sumner's folkways and mores, such assumptions bind not because they are strongly supported by evidence, but because they are unquestioned (Sumner 1906).

Perhaps a hundred years ago, when the social sciences were defining their measurement and analysis methodologies, the view that inherent differences in social and physical phenomena demanded different measurement and analysis methods may have seemed plausible, but no one with a deep understanding of modern physical science could entertain such a view today. Before the advent of modern science, we believed that physical objects were vague, evanescent, spontaneous and not governed by natural laws, just as we still think about cognitive phenomena. This, of course, is the essence of Plato's philosophy.

Plato's concept of knowledge was universal, absolute and unchanging. But the world of experience is particular, relative and constantly in flux. Plato concluded that there were two worlds, this imperfect, confusing and unknowable flux, and the World of Ideas. Science is anathema to Plato, useful only to carpenters and bricklayers; the true philosopher disengages from this world, hoping to remember or "lead out" of his memory (*e ducere* ("lead out from" in Latin, or *educate* in English) what could be recaptured from our past lives in the perfect World of Ideas, which is the only source of true knowledge.

Aristotle, Plato's student, disagreed, and held that this world consisted of two components: matter and form. Matter has no characteristics, but becomes "something" only when coupled with a universal, unchanging abstract form. It is the forms that are the basis of knowledge, but need to be abstracted from our material experiences by reasoning.

While these two philosophies differ in important ways, both deny that the world of experience provides the basis of knowledge. For Plato, all observations of this world should be avoided, since they only confuse the mind and hinder recollection of the "real" world of ideas. For Aristotle, observations are only a starting point, from which we must reason to the real basis of knowledge, abstract, unchanging form.

These two philosophies, Plato's through Augustine and Aristotle's through Aquinas, formed the philosophical foundation of Christianity, which, of course, believes that there are two worlds, the corrupt world in which we live, and the perfect world -- heaven or hell -- where we achieve perfect joy or perfect torment -- which comes after we leave this world.

At about the time of Galileo[3] and his successors, scientists increasingly began to abandon the idea that the world had an absolute character independent of our method of observation. Philosophers like DesCartes argued that measurement consisted not in making absolute statements about our experiences, but rather in making comparisons of observations to some standard. Scientists like Galileo put this comparative model into practice by comparing the rate of swinging chandeliers to a human pulse, measuring the degree to which a pea rises in slated water compared to the breadth of a finger, and measuring the distance a ball rolled down an inclined plane to the steady rhythm of a song.

Also abandoned was the concept of absolute, perfect, unchanging knowledge, which was replaced by the notion of imperfect theories which needed constant checking and revision based on observations.

Overwhelmingly, most social scientists never adopted this measurement strategy, but continue to believe that the world has an inherent character independent of the way we observe it. In fact, social scientists evaluate measurement systems in terms of the concept of *validity*, which, they believe, is the extent to which a measurement corresponds to the "true nature" of the thing being measured. Physical scientists, on the contrary, understand that what we observe is a function of the methods by which we observe, as Heisenberg pointed out:

> "(T)he idea of material objects that are completely independent of the manner in which we observe them proved to be nothing but an abstract extrapolation, something which has no counterpart in nature (Heisenberg, 1971, p. 85)."

This difference has had profound effects on the development of social science. Both physical scientists and social scientists understand that the "true nature" of the world is unavailable to them, but physical scientists abandon the quest for absolute truth, and focus on the relationships among the measurements they can make. Understanding that observations are dependent on "..the manner in which we observe...", physical scientists have striven to form international agreements about methods of observation, such as the Treaty of the Meter. In this way, science makes increasingly precise and agreed upon observations which lead to the social construction of a consensual reality *about our experiences of this world*. As Einstein says, "We are accustomed to regard as real those sense perceptions which are common to different individuals, and which are, in a measure, impersonal" (Einstein 1956).

The idea that knowledge is perfect and unchanging and absolute, however, is deeply ingrained in the global neural network that makes up the collective consciousness of social science, as is the idea that this world of experience is flawed and not the source of true knowledge. . Most social scientists continue to believe that observations made of this world of

---

[3] To be sure, even at the time of Socrates, Plato and Aristotle, some Greek scientists understood the comparative nature of scientific measurement, as did Aristarchus of Samos when he calculated the size of the earth and the moon and the distance to the sun. It is possible to show a direct link from these early scientists through Copernicus to Galileo and Newton.

experience are not the true object of our quest, but that, through an appropriate selection of assumptions, it is possible to "adjust" the measured values of observations to reveal the "true nature" of the variable measured. Psychometricians, for example, make assumptions about what the underlying geometry of a given configuration of data points "ought to be", and adjust measured values until they produce the "expected" geometry (Shepard 1962; Kruskal 1964). Proponents of Rasch measurement make assumptions about how a perfect scale ought to be distributed, and adjust measured values until those criteria are met (Rasch 1960).

Most social scientists, following Stevens, believe that different variables have inherent characteristics that affect their "measurability" (Stevens 1946). Some phenomena can only be categorized. Others can be arrayed in rank order, but without any magnitudes. Still others can be arrayed according to relative magnitude, but cannot be made into ratios, while still others can be expressed as ratios. Some psychometricians, such as Luce, have grave reservations about Stevens' model, but still believe in the existential character of variables and provide mathematical adjustments based on the distributional properties of observations to yield results that meet mathematical assumptions they believe would characterize "correct" measures (Luce 1997).

These differences in belief between physical and social scientists lead to irreconcilable differences in the definition of measurement. For physical science, the definition of measurement is clear and simple: measurement is *comparison to some standard.* For the social sciences, with their complex epistemological assumptions about the measurability of different phenomena, the definition is much more complex: measurement is *the assignment of numbers to observations according to some rule* (Stevens 1946).

And rules abound[4]. As a result, social science measurement procedure is very different from measurement in the physical sciences and engineering, and, in fact, by strict standards of scientific measurement, may not be be considered measurement at all. Among the most common

---

[4] Some rules are very complicated, as are the rules for a *genuine* Likert scale: first, gather a large population of statements about some object. Second, ask a sample of appropriately chosen people to place the statements (written on 4X6 cards) into piles (typically, eleven piles) so that similar statements go in the same piles. Third, record on the back of the card the number of the pile -- e.g., 1 through 11 -- in which the card was placed, and, fifth, repeat this for perhaps 100 respondents, then sixth, calculate the mean and standard deviation of each of the cards. Seventh, discard those cards that have high standard deviations, retaining only the 100 items with the lowest standard deviations. Eighth, assemble a questionnaire of the best 100 statements, which asks, for each statement, how much the respondent agrees with it with a five category checklist from strongly disagree through strongly agree, numbered 1 through 5. Ninth, multiply the score on the five point scale for each item by the mean of the item, and sum across all items. Tenth, normalize appropriately. The result is the attitude score for a given individual about that object. To make for even greater complication, many variations on this procedure exist.

rules is the simple numbered category rule, e.g., how favorable are you toward X, where 1= very unfavorable, 2= unfavorable 3= neutral, no opinion, 4 = favorable, 5 = very favorable.

While it is commonplace to note that reality is socially constructed, few generally understand that the Treaty of the Meter was the fundamental step in the process of constructing the concept of space as we understand it today. Before the Treaty, standard measures of time and distance were crude. In conducting his famous inclined plane experiment by which he arrived at the law of falling bodies, Galileo had no clock sufficiently precise to measure the duration of the fall, so he, first, used an inclined plan to slow the falling to a manageable rate, then -- as some now suspect -- sang or chanted, keeping time to the music, as he measured off the distance the ball rolled. Had he used the most common social science methods, however, the result would have been much different:

To test this assumption, Galileo's experiment was replicated with minor alterations:

A 48" long formica-topped table was tilted slightly by placing 200 page paper back books under two end legs. A 3/4" plastic ball was released at the higher edge of the table and allowed to roll down the incline until it dropped off the other edge. The time of rolling was measured with a Hewlett-Packard HP55 timer, and the distance the ball had rolled at one-second intervals was marked off. Instead of using a comparison to a standard rule, however, the distance rolled was measured using the most common social science rule: a numbered category scale, where 1-very short, 2=short, 3=neither long nor short, 4= far, and 5=very far. Ten trials were run for a total of 50 measurements. The mean distanced rolled was 2.76±..20. The power curve $s=at^b$ was estimated, with parameters $s=.96t^{.95}$, and fit with an r^2 of .908. Equally good (or bad) was the linear equation $s=a+b^t$, with parameters estimated as $s=-.02±.92^t$, which fit with an r^2 of .904. The worst fitting equation is the correct law, $s=a+bt^2$, whose parameters were estimated as $s=1.11+.149t^2$, and a fit of $r^2=.883$.

If Galileo's imitator had persisted, and constructed a new, bigger apparatus, and rolled more balls, the situation would have become even more confusing, since the numbers 1 through 5 in the first experiment would have referred to different distances than the same numbers would on the larger apparatus. If, for example, he built a table twice as long, "5" would mean roughly between 76 and 96 inches, whereas in the original study, "5" meant roughly 38-48 inches. Clearly, if physicists used the most common social science measurement rule, they would not have arrived at the same "laws of nature" we now know. Instead, their world would look contain an unpredictable component (the 10% unexplained variance in this experiment), and they might even have concluded that reality is not law governed, assigning the ball a measure of "free will."

It's not just the laws of nature that would be disrupted if physics were to adopt social science measurement procedures, but our entire picture of the world would be vastly different. Using the method of comparison to the standard meter -- the measurement rule for the physical sciences -- we are able to say that the diameter of the sun is 1.4 million kilometers, the diameter of the moon 3476 kilometers, the diameter of a US quarter 24.26 millimeters, and a US nickel

21.21 millimeters. We can say that the sun is 6.6^{10} larger than a nickel, about 66 billion times larger.

If we apply the most commons social science measurement rule, however, the world looks very different. On the face of it, the social reality that can be constructed out of five points scales is very crude, allowing for at maximum a ratio of five to one (i.e., something could be, at most five times larger than something else) compared to the 66 billion to one ratio of the sun to a nickel. In practice, the distortion is even worse. Consider the following experiment:

Twenty four undergraduate students were randomly divided into three groups. Each answered four questions. Each question was collected before the next was administered. The questions were: How large is the sun? How large is the moon? How large is a quarter? How large is a nickel? Responses were 1=very small, 2=small, 3=neither large nor small, 4=large, 5=very large. The questions were administered in this order:

Group 1	Group 2	Group 3
Moon	Sun	Nickel
Sun	Moon	Quarter
Nickel	Nickel	Moon
Quarter	Quarter	Sun

Table 1: Order of presentation for three groups

The results of these measurements are:

	Group 1	Group 2	Group 3
Moon	3.38±1.41	4.00±.93	3.88±1.25
Sun	4.25±1.16	4.25±1.04	4.13±.99
Nickel	2.13±.64	2.25±.71	2.50±.53
Quarter	2.13±.64	2.38±.52	3.25±.71

Table 2: Size of Moon, Sun, Nickel and Quarter for three groups

Although the actual ratio of the size of the sun to the size of a nickel is about 66 billion to one, and the maximum theoretical ratio possible in a five point scale is five, the actual ratios measured in the experiment are shown in the following table:

	Group 1	Group 2	Group 3
Moon/Sun	.80±1.82	.94±1.40	.94±.59
Moon/Nickel	2.0±1.72	1.89±1.07	1.65±1.36
Sun/Quarter	2.0±1.32	1.79±1.16	1.27±1.21
Nickel/Quarter	1.0±1.17	.95±.88	.77±.89

Table 3: Ratios of sizes of Moon, Sun, Nickel and Quarter for three groups

The largest ratio actually measured in Table 3 is 2.0, showing respondents estimated the Moon to be twice the size of a nickel, and the Sun to be twice as large as a quarter. Respondents report that the moon is between 80 and 94% of the size of the sun, and one group reports the sun is only 27% larger than a quarter.

Two Worlds Again? The Sample and the Population

Once again, however, the deeply ingrained notion of two worlds -- the world of unreliable experience and the real or genuine world -- arises to worsen the situation. In statistics, social scientists are taught that their observations are always constrained to samples drawn from a universe or population. Although the sample is the only thing we can observe, our real interest is in the (unobservable) population. We can, however, on the basis of some mathematical and statistical assumptions, reason to what the "real" population must be like. In the present case, using the standard rules of inference from sample to population, if standard 90 or 95 percent confidence intervals are used, we would conclude that these subjects show no significant differences in size between the Sun, the Moon, a quarter and a nickel. In the platonic "population", our research shows that the Sun, the Moon, a quarter and a nickel are exactly the same size.

The point of this exercise is not to show the low level of education of the typical undergraduate. These students may not know that the Sun is 66 billion times larger than a nickel, but they do know that it is vastly larger. The standard social science methodology, however, *prevents them from expressing this knowledge* (Cairns 2007).

In the physical sciences, if theory doesn't agree with observation, the theory is wrong. But not in social science -- or medical science, either. In the largest health related study of all time, researchers at Harvard University found beneficial effects of dietary fat reduction on heart disease, breast cancer and colon cancer in their 25 year sample of 161,000 women, but the result was not statistically significant, so they were "forced" to conclude that there was no beneficial

effect in the population. The consternation in the investigator's words are apparent in this transcript from Talk of the Nation:

> The federally funded study, known as the Women's Health Initiative, tracked 161,000 women, lasted 15 years, and cost $725 million. It was designed to test different strategies for preventing heart disease, bone thinning osteoporosis, and breast and colorectal cancer. It made news in 2002 when its findings created doubt that hormone therapy was good for women's overall health. It again made news last month when final results did not conclusively show that a low-fat diet reduces the risk for heart disease, breast cancer, and colon cancer. It also did not show whether or not calcium and vitamin D offer protection against broken bones.

> CLARK:

> This morning I was listening to Diane Rehm Show and the doctors, previous leaders of the NIH, seem to say, if I got my information correct, that calcium along with vitamin D helped reduce hip fractures by twenty some odd, 26 to 29 percent. And then at the beginning of your show here, one of your comments to introduce the show was the saying that calcium, the study showed calcium may not matter, or didn't help. And if I had just heard that and then turned my radio off, I might call my mom and tell her don't worry about her taking the calcium.

> Dr. HOWARD: Well, I'm going to defend the press there in that the distinction here is that, as trialists, there is a scientific way that one has to present the result. And first, you have to show the group as a whole, and the reduction in hip fractures in all the group, all the ages, was not, it didn't reach that level of statistical significance. It was close again, but it didn't make it. So, we always have to say that first. And that's of course what the headlines picked up. But what he's pointing out is, as soon as you look to the next level you see for all the women over 60, there was a 21 percent reduction. And for the women who actually took their pills, because whenever you do a trial, and we had quite good adherence in that trial, in other words people taking their pills, but some didn't, if you took them out, then there was a 29 percent reduction in the whole group. So that's the problem. We have to report it, in a scientific article, the way it's expected for good science, and that then the news report is going to show the primary finding. And, if people don't read past the headline, they don't get the rest of it.

Dr. Howard's disclaimer on National Public Radio did little to ameliorate the public confusion when statistical significance testing compelled the researchers to disregard the obvious benefits of low fat diets on health and report that the results were not significant, and that, in the population, they couldn't reject the hypothesis that there was no relationship between dietary fat

consumption and cancer and heart disease. Clearly the theory -- for statistical inference theory is indeed a theory -- outweighs the observations outside the physical sciences[5].

Who Needs Measurement, Anyway? The Correlation Coefficient

While in physical science, measurement always means comparison to a standard, in the social sciences, measurement means assigning numbers to observation according to some rule, and their are thousands of alternative rules. The fact that measurements made by different investigators in different areas for different topics at different times based on different rules were incommensurate was widely understood early on, and led to the formation of a committee established in 1932 by the British Association for the Advancement of Science to investigate the possibility of genuine scientific measurement in the psychological and behavioral sciences. This committee, which was known as the *Ferguson committee*, published a Final Report (Ferguson 1940) which was highly critical of measurement in the social and behavioral sciences. One of the results of the report was Stevens' model of measurement already discussed (Wikipedia 2010). But even earlier, Karl Pearson proposed an alternative method which, he believed, eliminated scales of measurement altogether.

Pearson was convinced that the only thing of interest to science was the degree to which variables covaried. By expressing measured values as deviations from their mean values and multiplying paired values together, Pearson created a sum of products that would be large if positive deviations on one variable co-occurred frequently with deviations on another variable, large and negative if positive values on one variable co-occurred frequently with negative values on another variable, and zero if they covaried randomly. Of course the sum of products would depend on the sample size, so Pearson used the average of the cross products instead of the sum.

But the size of the average of the cross products would also depend on the scale of the variables -- variables scaled on 10 point scales would count more than variables scaled on a five

---

[5] R. A. Fisher is usually considered the founder of mathematical statistics, but the standard form of inferential statistics in wide use today follows the work of Jerzy Neyman and Egan Pearson (Karl Pearson's son), which Fisher himself opposed (see Klamer, A., McCloskey, Donals N., Solow, Robert M. (1988). The Consequences of Economic Rhetoric. Cambridge, Cambridge University Press.

One of the main vehicles for the dissemination of the inferential statistical model was Biometrika, a journal founded in 1901 and edited by Karl Pearson until his death in 1936, then edited by his son Egan until 1966. The spread of these methods through the social sciences was facilitated by the foundation of Econometrica in 1933 and Psychometrika in 1936. Although represented as orthodoxy in virtually all social science methods textbooks, it has been controversial from its inception and the past several decades have seen increasing attacks (Levine, T. R., Weber, Rene, Hullett, Craig, Park, Hee Sun, Massi Lindsey, Lisa L. (2008). "A Critical Assessment of Nul Hyothesis Significant Testing in Quantitatice Communication Research." Human Communication Research **34**(2): 16.
, Hurlbert, S. H., Lombardi, Celia M. (2009). "Final Collapse of the Neyman-Pearson decision theoretic framework and rise of the neoFisherian." Ann. Zool. Fennici **46**: 38.
.

point scale. So Pearson divided the average cross products by the lengths of the variables, assigning all variables a unit length. The resulting scalar product is equal to the product of the variable lengths and the cosine of the angle between the variable vectors. Since the variable vectors have been normalized to unit length, the resulting Pearson Product Moment Correlation -- r -- is actually the cosine of the angel between the two variable vectors.

The Pearson Product Moment Correlation, or simply correlation coefficient, is widely used as a basis for measuring the relationships among variables in the social sciences. But is it possible to construct a meaningful scientific world based solely on the cosines of angles? Is a world without magnitudes -- or a world where every magnitude equals one -- useful for scientific understanding?

In fact, the level of distortion introduced by Pearson's procedure is immense -- far to large to provide a useful basis for serious observation. Figures 1 and 2 provide a good demonstration of the level of distortion introduced the the Pearson Product Moment Correlation: Both figures are based on a matrix of intercity distances, corrected for the curvature of the earth. Figure 1 shows the principle axes of the centroid scalar products derived from these distances. It is a very accurate map showing the precise locations of all the cities.

Figure 1: Principle axes of intercity distances

Figure 2, on the other hand, is based on the principle axes of the correlation matrix derived from the distances. I shows the distortion inherent in all correlational analyses, where, due to the normalization rule, all objects reside on the surface of a unit hypersphere.
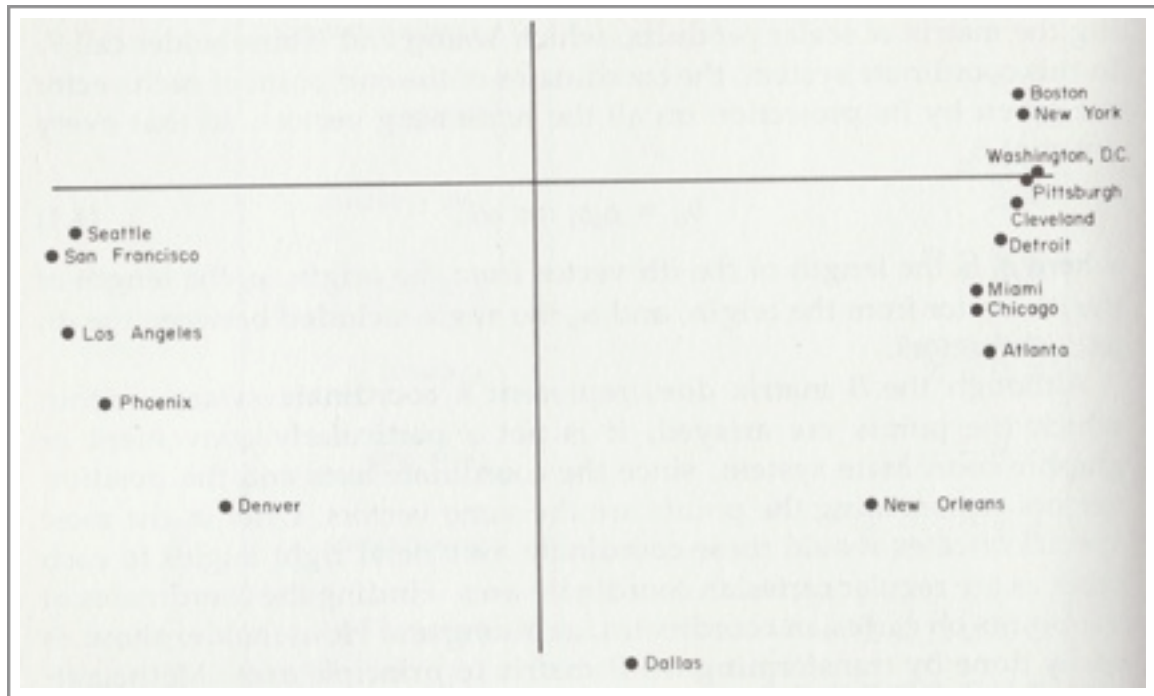

Figure 2: Principle axes of intercorrelations among intercity distances

Time

Although time in social science can be measured exactly the same as in physical science, so deeply engrained into the social science neural network is the notion of inherent vagueness and imprecision that it seldom is. Often time is categorized as pre-post, time one, time two and time three, or not at all.

One of the classic topic areas for Communication researchers, for example, is persuasion and attitude change (Dinauer 2003). Most theories of attitude change assume explicitly or implicitly that attitudes are dynamic equilibrium systems, which tend to return toward an equilibrium position when perturbed (the alternative is thought to be too unstable to form the basis of a viable society.)

Whether this theory is true or not (no theory has ever been decisively rejected yet), it implies a model in which attitudes, when perturbed from equilibrium, would move away from the equilibrium position, and return over time. Depending on the degree of damping, friction and restoring force in the system, the attitude might overshoot it's original equilibrium position and

even oscillate around that position over time, just as a weight on a spring would oscillate around its equilibrium position when disturbed.

Hundreds of attitude change experiments, however, in which different methods of perturbing the system are tried and results "measured" involve simple random assignment of cases to treatment or control conditions, with posttest-only measures -- generally considered a sound experimental model by social scientists. The problem is, of course, when one is measuring the position of an oscillating system, *when* the observations are taken is crucial. Depending on how long after the perturbation the observation is made, one will observe a movement in the direction of the perturbing force, no movement, or even a movement in the opposite direction -- which is exactly what the research record shows. Attitude change experiments in which time is carefully measured exist, but are rare (Fink 2009)(Barnett, Chang et al. 1991; Fink, Kaplowitz et al. 2002; Fink, Cai et al. 2003; Dinauer and Fink 2005; Chung and Fink 2008; Chung, Fink et al. 2008).

Lessons and recommendations

Around a century ago, some philosophers and social scientists, believing there were inherent differences in the subject matter of the social and physical sciences, proposed alternatives to the methods in general use in physical science for the social sciences. The purpose of the present article is to suggest that the original notion of a fundamental, a priori difference between the subject matter of physical and social science was incorrect. Moreover, the alternative methods adopted instead have insulated the social sciences from the effects of observation.

Since reality is socially constructed, we can inquire as to the social reality constructed by social science methodology, and find it unsatisfactory and unscientific. In fact, there is no reason the methods of physical science and engineering cannot be applied directly to social, cognitive and cultural phenomena. There is a small cadre of social scientists who reject the notion that there are two kinds of science, and who have applied the same principles of measurement and analysis to social, cognitive and cultural phenomena. Those who have done so have met with considerable success.

Measurement

The most primitive variables underlying physical processes are distance and time. Similarly, for cognitive and cultural processes, the primitive variables are the "distance" or difference between perceived stimuli and time. Measurement of perceived distance or difference consists in comparisons to a standard distance or difference. Woelfel and Fink suggested a simple rule where the distances between pairs of stimuli are compared to some standard or "criterion" distance in the form "If i and j are u units apart, how far apart are x and y?" (Woelfel and Fink

1980) A more recent version of this rule is given in figure 3, which uses sliders to estimate the dissimilarity among pairs of object visually and numerically:

Figure 3: Typical instructions for "sliders"

This measurement strategy corresponds exactly to physical practice, and produces precise and useful measurements. Changing the criterion pair and changing the modulus assigned to the pair produces changes in the size of the numbers reported, but the structures remain the same (Gordon 1976; Gordon and Deleo 1976, April; Bass, Gordon et al. 2008; Woelfel and Evans 2009).
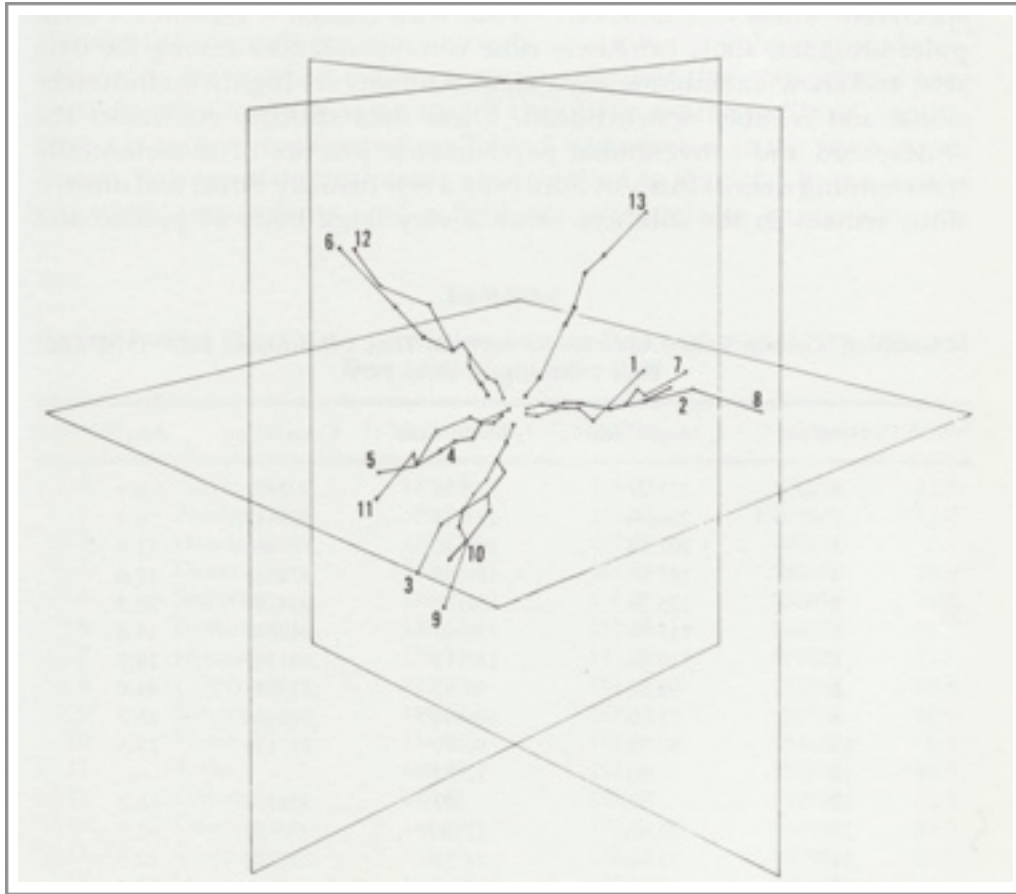
Figure 4: Comparison of Galileo spaces using different criterion pairs
and different moduli in order of size

As of April 30, 2010, the Galileo website lists 147 papers, books, theses, dissertations and other documents based on this system of measurement, and none of them report a failure to produce precise, useful results (Woelfel 2010)

Analysis

Observation trumps theory, and observed values must not be transformed away due to theoretical considerations or based on philosophical or mathematical assumptions. Transformations with no inverse are impermissible. It's perfectly alright to take a log transform to trim long straggly distributions and clear up the middle, because you can always go back. But it is never permissible to transform any variable in a way that cannot be transformed back. Conversion of measured values into standardized scores is never a good idea.

Transformation to principle axes

Measurement of pairwise dissimilarities usually yields square, symmetric matrices of distances. Transforming the scalar products of these matrices to principle axes provides a convenient coordinate system in which cognitive and cultural processes can be observed as motions of points through the space over time. Transformations that modify the raw dissimilarities to make the space fit preconceived geometrical or other assumptions should not be allowed. In general, the spaces resulting from this transformation are high dimensional and Riemannian (Woelfel and Fink 1980).

Measuring Attributes

Attributes are never lines or line segments, such as "good-bad" or "happy-sad." "Bad" is not the opposite of "good", nor is "sad" the opposite of "happy." All attributes should be considered points, and one can move closer to or further from them. There is no limit to how far away from "happy" one can be. There is no reason any person should lie on a line connecting "happy" and "sad," and, indeed, this has never been observed by any social scientist and would be very, very unlikely in a statistical sense. It is possible to move closer to "good" and "bad" at the same time, as it is possible to mover further from both "good" and "bad" at the same time.
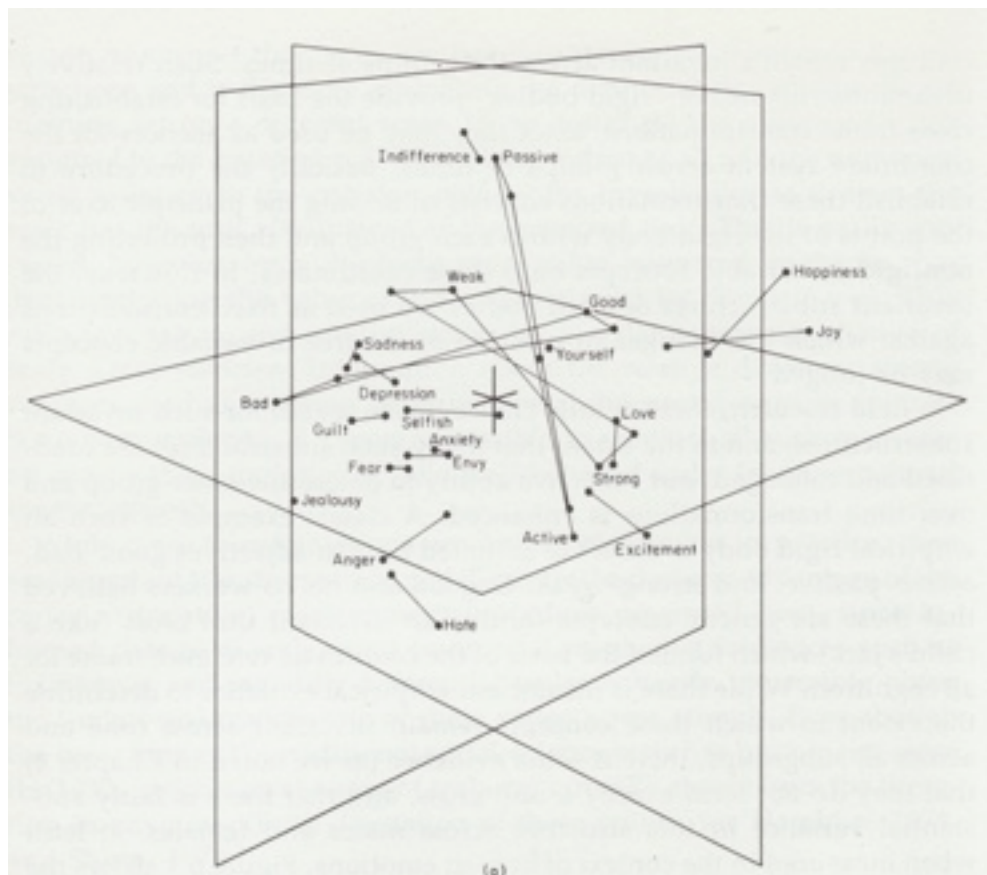

Figure 5: Galileo plot of Semantic Differential attributes
(adapted from Woelfel & Fink, 1980)

Figure 5 shows the first three dimensions of the major terms in Osgood's semantic differential theory, good-bad, active-passive and strong-weak. Osgood assumes these are independent of each other, i.e., mutually perpendicular, that they all pass through the origin, which is a point of meaninglessness, that they are all the same length, and that the length from one side to the origin is the same as the length from the other (that is, from good to neutral is the same distance as from bad to neutral). Since this is meant to be universally true, the pattern will be the same for males and females. As Figure 5 shows, none of this is true. None of the "attributes" pass through the origin, which is nothing but the center of this particular subspace (Woelfel and Fink 1980).

Figure 6 shows the relationships among six fictitious people and three of their attributes: height, political orientation and intelligence. Figure 6 shows that the categories very tall, tall short and very short do not lie on a straight line segment. Very intelligent, intelligent, unintelligent and unintelligent do not lie on a straight line segment, nor do very conservative, conservative, liberal and very liberal. None of the intervals are the same, and none of the attributes intersect any of the others.
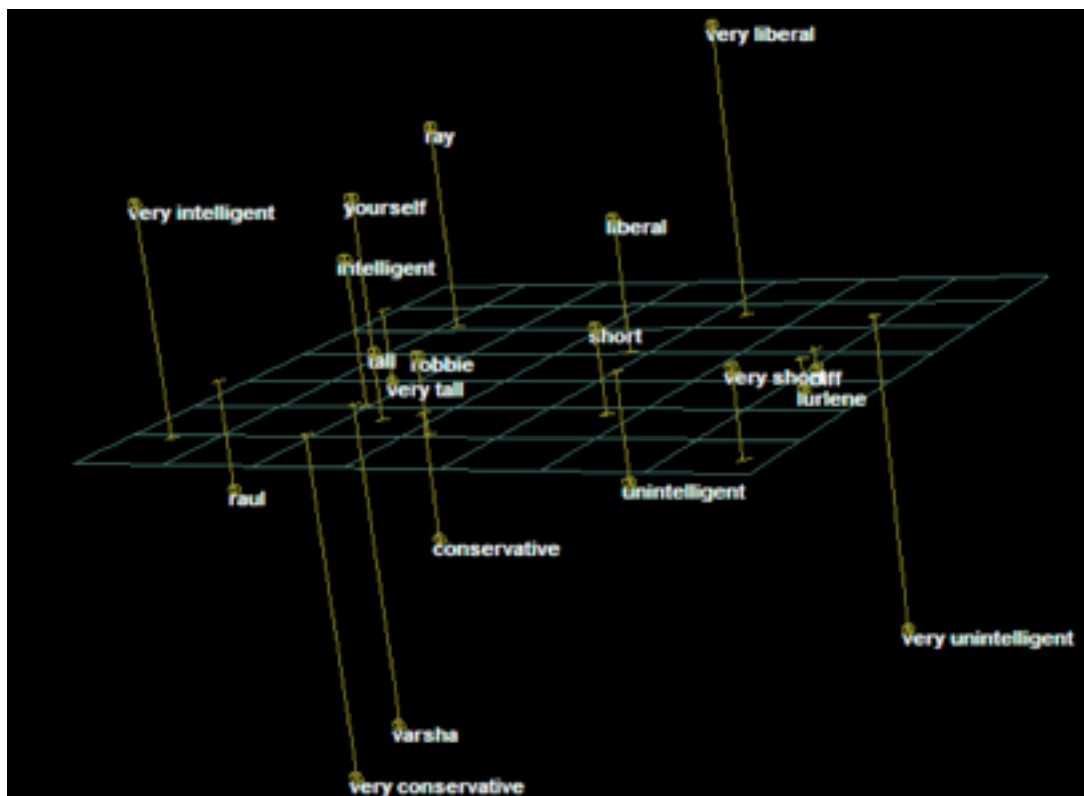

Figure 6: Plot of six fictitious people and three of their attributes

Summary and Conclusions

Physical science has only one definition for measurement, *comparison to a standard*. Social science, on the contrary, defines measurement as *assignment of numbers to observations according to some rule*. Hundreds, perhaps thousands, of such rules have been put forward, and none have been rejected. Observations are not independent of the measurement process, and so observations made by social scientists produce vastly different results that those of physical science. The assumption that these differences are consequences of the inherent character of the variables measured is unsupported by any data, and, in fact, the application of social science measurement rules to simple, well understood physical phenomena yields absurd results.

If the social sciences are to develop a body of scientific theory similar to that common to physical sciences, it will be necessary to adopt the same measurement rule. Evidence accumulated over many decades by a small, independent group of researchers has shown that the comparative rule produces precise, high quality data very comparable to those obtained in physical science research.

The essence of science is ultimate reliance on observation, but standard social science methodology requires that observations yield to theory and philosophical/mathematical assumptions. In particular, the use of inferential statistics leads investigators to ignore or disregard actual observations based on statistical assumptions and theory. A large literature critical of statistical hypothesis testing has developed, but perhaps the best advice is to absolve investigators from making any decision at all about what they have observed, and simply require that they report what they have found, and the levels of precision around those findings, leaving decision making to posterity after sufficient replication or lack thereof by other investigators -- as is the practice of physical science.

Indeed, a common error to which neural networks are prone is to misidentify a pattern early on, which leads to an increasing inability to avoid repeating the same error in the future (Woelfel 2010). The collective decision of the social sciences to adopt S. S. Stevens fourfold classification of scale types, Karl Pearson's correlation coefficient and Jerzy Neyman and Egan Pearson's test against the null hypothesis as orthodox social science method have been extremely costly mistakes.

Bibliography

Barnett, G. A., H. Chang, et al. (1991). "Seasonality in TV viewing: A mathematical model of cultural processes." <u>Communication Research</u> **18**(6): 755-772.

Bass, S., T. Gordon, et al. (2008). "Mapping perceptions related to acceptance of smallpox vaccination by hospital emergency room personnel." <u>Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science</u> **6**(2): 179-190.

Cairns, P. (2007). <u>HCI...not as it should be: Inferential Statistics in HCI research</u>. BCS-HCI '07 Proceedings of the 21st British CHI Group Annual Conference on HCI 2007: People and Computers XXI ...but not as we know it - ACM Digital Library.

Chung, S. and E. L. Fink (2008). "The cognitive dynamics of beliefs: The effect of information on message processing." <u>Human Communication Research</u> **34**: 477-504.

Chung, S., E. L. Fink, et al. (2008). "The comparative statics and dynamics of beliefs: The effect of message discrepancy and source credibility." <u>Communication Monographs</u> **75**(2): 139-170.

Dinauer, L. (2003). Attitude and belief change in explicit and implicit concept hierarchies: A comparison of two models of interattitudinal structure. College Park, MD, University of Maryland. **PhD**.

Dinauer, L. and E. L. Fink (2005). "Interattitude structure and attitude dynamics: A comparison of the hierarchical and galileo spatial-linkage models." <u>Human Communication Research</u> **31**(1): 1-32.

Einstein, A. (1956). <u>The Meaning of Relativity</u>. Princeton, Princeton University Press.

Ferguson, A., Myers, C.S., Bartlett, R. J., Banister, H., Bartlett, F. C., Brwn, W., Campbell, N. R., Craik, K. J. W., Dreyer, J., Guild, J., Houstoun, R. A., Irwin, J. O., Kaye, G. W. C., Philpott, S. J. F., Richardson, L. F., Shaxby, J. H., Smith, T., Thouless, R. H., Tucker, W. S. (1940). "Quantitative estimates of sensory events: Final report of the committee appointed to consider and report upon the possibility of quantitative estiates of sensory events." <u>Advancement of Science</u> **1**: 18.

Fink, E. L. (2009). "The FAQs on data transformation." <u>Communication Monographs</u> **76**(4): 379-397.

Fink, E. L., D. Cai, et al. (2003). "The semantics of social influence: Threats versus persuasion." <u>Communication Monographs</u> **70**: 295-316.

Fink, E. L., S. A. Kaplowitz, et al. (2002). Oscillation in beliefs and decisions. The persuasion handbook: Theory and practice J. P. Dillard and M. Pfau. Thousand Oaks, CA, Sage Publications: 17-37.

Gordon, T. (1976). Subject abilities to use MDS: Effects of varying the criterion pair, Retrieved from Galileoco, Galileo Literature #21. (http://www.galileoco.com/CEtestLit/literature.asp)

Gordon, T. and H. C. Deleo (1976, April). Structural variation in "Galileo" space: Effects of varying the criterion pair in metric multidimensional scaling. International Communication Association conference. Portland, OR.

Haller, A. (2000). Societal stratification. Encyclopedia of sociology. E. Borgatta and R. Montgomery. New York, Macmillan Reference, USA. 4.

Hurlbert, S. H., Lombardi, Celia M. (2009). "Final Collapse of the Neyman-Pearson decision theoretic framework and rise of the neoFisherian." Ann. Zool. Fennici 46: 38.

Klamer, A., McCloskey, Donals N., Solow, Robert M. (1988). The Consequences of Economic Rhetoric. Cambridge, Cambridge University Press.

Kruskal, J. B. (1964). "Nonmetric multidimensional scaling: A numerical method. ." Psychometrika 29: 14.

Levine, T. R., Weber, Rene, Hullett, Craig, Park, Hee Sun, Massi Lindsey, Lisa L. (2008). "A Critical Assessment of Nul Hyothesis Significant Testing in Quantitatice Communication Research." Human Communication Research 34(2): 16.

Luce, R. D. (1997). "Quantification and symmetry: ommentary on Michell 'Quantitatie science and the definition of measurement in psychology'." British Journal of Psychology 88: 4.

Rasch, G. (1960). Probabalisitc models for some intelligence and attainment tests. . Copenhagen, Danish Institute for Educational Research.

Shepard, R. N. (1962). "The analysis of proximities: Multidimensional scaling with an unkown distance function." Psychometrika 27(2): 15.

Stevens, S. S. (1946). "On the Theory of Scales of Measurement." Science 103(2684): 4.

Sumner, W. G. (1906). Folkways. Boston, Ginn and Co.

Wikepedia (2010, 21 April 2010). "Level of measurement." from http://en.wikipedia.org/wiki/Level_of_measurement.

Woelfel , J. (2010). Social Networks as the Substrate of Cultural Processes. Honolulu, HI, University of Hawaii under a grant from the Air Force Office of Scientific Research.

Woelfel, J. and C. Evans (2009). <u>Galileo and multidimensional scaling</u>. Amherst, NY, RAH Press.

Woelfel, J. and E. L. Fink (1980). <u>The measurement of communication processes: Galileo theory and method</u>. New York, Academic Press.

Woelfel, J. E., C. (2010). "Galileo Literature." from http://galileoco.com/CEtestlit/literature.asp.