How Measurement Works

William Bricken
December 2007

One fundamental difference between Science and Art is that Science relies upon numbers in general and measurement in particular. There are several different kinds of numbers, just like there are several different types of Art. The concreteness of numbers, their connection to physical reality, makes Science different than Art.

The meaningful use of numbers relies upon making a measurement. *Measurement* is an act of comparison, an object to be measured is compared to marks on a standardized measurement tool. *What a number means* differs significantly when what is being measured is physical (e.g., measuring distance using a ruler) or non-physical (e.g., measuring knowledge using a multiple choice test).

The simplest use of numbers for measurement is *tally arithmetic*, the idea of making a single mark that corresponds to a particular object. This is the way that physical items have been tallied, or counted, for ten millennia. The collection of marks illustrates the number of objects, even after the physical collection has been disbanded.

●●●● + ●●●●●●● = ●●●●●●●●●●●

A closely related idea is to climb a counting ladder, one step at a time. We give names to each step (*two*, *three*, *four*…) and thus construct shorthand labels that permit convenient reference to a multitude of tally marks.

$$4 + 7 = 11$$

Numbers are used to describe physical attributes. We first define a *standard unit*, such as a meter, or a pound, or a second. Then we construct a *measuring tool* by calibrating a physical object using the standard unit. Measuring tools construct automatic tallies using standard units (i.e. they count). *Every meaningful number* is read from a tool that is calibrated by some type of standardized unit. Analogously, in the Art of written language, every idea is read from words that are defined by some type of intended meaning. The difference is that words do not require a physical referent (they can refer to concepts) while numbers do require a physical referent (they cannot refer to concepts). Thoughts, for example, are notoriously difficult to count. The idea of "one-and-a-half concepts" does not make sense.

Counting specific objects is *precise*. However, a degree of imprecision is introduced into the idea of measurement when we try to read between the unit marks of a measurement tool. This imprecision converts integer numbers (tallies) into *statistical numbers* (probabilities). Statistics puts meaning into the numbers that result from inexact measurement by positing that there is, in fact, a True measure, but that the numbers we read from the measurement instrument are fuzzy. Given a True measure of 50, a fuzzy numerical measurement may at one time read "48", but at another time read "56". Actually, all measurement tools yield fuzzy numbers, either at a fine grain of detail, or due to physical disturbance and/or human error during measurement.

We can never know a True measure by looking at a statistical number, all we can say is that an observed measurement is close to the hypothetical True measure, within some degree of accuracy. The wider the range of ambiguity around an observed measurement, the higher our level of confidence is that the True measure lies within that range.

```
                True      observed
                  ↓          ↓
-----------{-----------------------------------------}-----------
           /                                         \
```

range of confidence around an observed measurement

All statistical numbers within the range of confidence have the same meaning since they refer to the same True measure. For purposes of measurement, the observed number 54 (illustrated below), with an accuracy of ± 8 units, has the same meaning as the numbers 46 and 62, and all the other numbers in between. The main idea of statistical numbers is that *within the range of confidence, different numbers mean the same thing*.

```
               True=50    observed=54
                  ↓          ↓
-----------{-----------------------------------------}-----------
           /                                         \
```

46 < *range of meaning of "54"* > 62

Analogously, the meaning we read into a written word is only somewhat close to its intended meaning. Fuzzy words permit, for example, "cat" and "feline" to convey the same idea. Taking a statistical number to *be* the True measure is analogous to saying that every word has one and only one exact meaning.

Statistical numbers have been applied to the measurement of non-physical things, such as a student's knowledge. Measurement of non-physical properties is always *approximate* for three distinct reasons:

1) *Ambiguity*: We can never know what the True measure is, or for that matter, whether or not a True measure exists, because there is no concrete Real thing to define what True means.

2) *Reliability*: All non-physical measurements are fuzzy because calibrating a measurement instrument is very difficult when there is no physical referent.

3) *Validity*: The standard unit itself is inexact, can vary between measurements, and is usually poorly defined both technically (*measurement validity*) and conceptually (*content validity*).

Hopefully, our decisions are guided by True measures; hopefully we seek to avoid mistaking a finger pointing at the moon for the Moon itself. Scores on the widely used COMPASS placement test, for example, have a published range of accuracy of about ± 8 points on either side of an observed score. A student with an observed score of 54 most likely has a True score between 46 and 62. But to say that the particular score of 54 differs from a score of 46 or a score of 62 is to call upon an Artistic rather than a Scientific use of numbers.