Abstracting Some Aspects

Measure what is measurable, and make measurable what is not so. Galileo Galilei

The process of measurement begins long before any data are collected. The starting point is a notion, or even better, a theory about an aspect of a class of things we want to understand better, maybe even do some science on. Successful measurement depends on clear thinking about the aspect and clever ideas for the agents. This is much more challenging and much more rewarding than any mathematical gymnastics that might be performed to fit model to data.

All analogies are limited but some are useful. Considering aspects of things far removed from cognitive traits may help avoid some of the pitfalls encountered when working too close to home. *Hardness* is a property of materials that is hard to define but we all know what it is when it hits us. *Color* is a narrow region of a continuous spectrum that non-physicists tend to think about as discrete categories. *Temperature* is an intimate part of our daily lives, which we are quite adept at sensing and more recently at measuring, but the closely connected idea, *heat*, may actually be more real, less bound to conventions and populations. If I could scale the proficiency of professional football teams and reliably predict the outcomes of games, I wouldn't be writing this.

Hard Headed: the importance of being sufficient

We will start far outside the cognitive world with the very ordinary idea of *hardness*, the quality of being firm or solid. This description, while completely reasonable and perhaps meaningful, does not say much about how one might measure it. Attempts to quantify the idea of hardness led to more operational definitions: e.g., the degree to which the surface of a material may be scratched, indented, abraded, or machined or, conversely, the ability to resist being scratched, indented, abraded, or machined.

Everyone knows that you test if your new diamond is genuine by scratching a mirror; diamond is harder than glass. This notion of hardness produced the *Mohs* scale, which relies on ten materials (Table II.1) that are used to determine what scratches what. A material (e.g. garnet) that scratches quartz and is scratched by topaz has a Mohs scale value of 7 but we can't distinguish a just-barely-over-7 material from an almost-to-8. The Mohs' scale is useful for separating diamonds from glass but not precise enough to differentiate grades of steel.

While not a measure, the Mohs score is a very sufficient statistic.

Table II.1: The Ten Substances in Mohs Scale of Hardness

Scale of Hardness		
Substance	Mohs	Shore Units
Talc	1	1
Gypsum	2	2
Calcite	3	9
Fluorite	4	21
Apatite	5	48
Orthoclase	6	72
Quartz	7	100
Topaz	8	200
Corundum	9	400
Diamond	10	1500

This process is certainly general and reproducible but we don't know if this is an interval measurement scale or not. Because diamond always scratches corundum, which always scratches topaz, etc., we have no way to determine the spacing. In the absence of evidence to the contrary, we have to assume these are just ordinary ordered categories, not measures.

Students of earlier approaches to scaling might recognize this as a *Guttmann* scale, which is tricky to implement and restricted in its scaling, but has one very profound property: *a Guttmann-Scale score completely captures all observed responses and all possible future responses.* With a Mohs-Scale score of 7, we know exactly which materials our rock scratched and which it did not and which materials, presented or not, it could and could not scratch. While not a measure, the Mohs score is a completely sufficient statistic.

Later efforts (e.g., Brinell, Rockwell, Vickers) used various devices for slamming different kinds of hammers into the targets. All these techniques report values, something like the depth of the indentation, the height of the rebound, or the energy lost in the process, that are specific to the load, shape, and duration. The values from any of these methods and scales could be compared but only to those obtained by the same method under the same conditions, i.e., they took the same form of the test.

There are several Brinell scales and more than a dozen Rockwell scales with different loads and shapes of hammers, intended for materials with different properties and in different ranges of hardness. There is nothing inherently wrong with having different agents appropriate to different objects if we have some way to compare across agents. This is like having different *rulers* for measuring the sizes of molecules, widths of rooms, waists of people, and diameters of galaxies.

After much banging of heads and rocks, Shore units eventually emerged, which could be used for the Mohs materials (see Table II.1) or anything else we might encounter. This is an attempt to free the measurement from the specifics of the situation by applying some basic Newtonian physics to control or eliminate effects due to the mechanics of the device. The result was a scale that can be called measurement.

The issue of dimensionality does not even come up in this discussion. The topic is hardness, not to be confused or confounded with brittleness, density, tensile strength, mass, luster, color, or any other aspect that these materials might share that might be interesting to someone else or at some other time. But if anyone ever encounters a substance that scratches topaz (*Mohs 8*) and is scratched by quartz (*Mohs 7*), everyone will be back to the drawing board.