I started this post in Hong Kong airport, having just finished one conference and heading to Innsbruck for another. The Hong Kong meeting was on psychometrics and the Innsbruck conference is on imprecise probabilities (believe it or not, these topics actually do overlap). Anyhow, Annemarie Zand Scholten gave a neat paper at the math psych meeting in which she pointed out that, contrary to a strong intuition that most of us have, introducing and accounting for measurement error can actually sharpen up measurement. Briefly, the key idea is that an earlier "error-free" measurement model of, say, human comparisons between pairs of objects on some dimensional characteristic (e.g., length) could only enable researchers to recover the order of object length but not any quantitative information about how much longer people were perceiving one object to be than another.

I'll paraphrase (and amend slightly) one of Annemarie's illustrations of her thesis, to build intuition about how her argument works. In our perception lab, we present subjects with pairs of lines and ask them to tell us which line they think is the longer. One subject, Hawkeye Harriet, perfectly picks the longer of the two lines every time—regardless of how much longer one is than the other. Myopic Myra, on the other hand, has imperfect visual discrimination and thus sometimes gets it wrong. But she's less likely to choose the wrong line if the two lines' lengths considerably differ from one another. In short, Myra's success-rate is positively correlated with the difference between the two line-lengths whereas Harriet's uniformly 100% success rate clearly is not.

Is there a way that Myra's success- and error-rates could tell us exactly how long each object is, relative to the others? Yes. Let p_{ij} be the probability that Myra picks the i^{th} object as longer than the j^{th} object, and $p_{ji} = 1 – p_{ij}$ be the probability that Myra picks the j^{th} object as longer than the i^{th} object. If the i^{th} object has length L_i and the j^{th} object has length L_j, then if $p_{ij}/p_{ji} = L_i/L_j$, Myra's choice-rates perfectly mimic the ratio of the i^{th} and j^{th} objects' lengths. This neat relationship owes its nature to the fact that a characteristic such as length has an absolute zero, so we can meaningfully compare lengths by taking ratios.

How about temperature? This is slightly trickier, because if we're using a popular scale such as Celsius or Fahrenheit then the zero-point of the scale isn't absolute in the sense that length has an absolute zero (i.e., you can have Celsius and Fahrenheit readings below zero, and each scale's zero-point differs from the other). Thus, 60 degrees Fahrenheit is not twice as warm as 30 degrees Fahrenheit. However, the differences between temperatures can be compared via ratios. For instance, 40 degrees F is twice as far from 20 degrees F as 10 degrees F is.

We just need a common "reference" object against which to compare each of the others. Suppose we're asking Myra to choose which of a pair of objects is the warmer. Assuming that Myra's choices are transitive, there will be an object she chooses less often than any of the others in all of the paired comparisons. Let's refer to that object as the Jth object. Now suppose the i^{th} object has temperature T_i,the j^{th} object has temperature T_j, and the J^{th} object has temperature T_J which is lower than both T_i and T_j. Then if Myra's choice-rate ratio $p_{iJ}/p_{jJ} = (T_i – T_J)/(T_j – T_J)$, she functions as a perfect measuring instrument for temperature comparisons between the i^{th} and j^{th} objects. Again, Hawkeye Harriet's choice-rates will be $p_{iJ} = 1$ and $p_{jJ} = 1$ no matter what T_i and T_j are, so her ratio always is 1.

If we didn't know what the ratios of those lengths or temperature differences were, Myra would be a much better measuring instrument than Harriet even though Harriet never makes mistakes. Are there such situations? Yes, especially when it comes to measuring mental or psychological characteristics for which we have no direct access, such as subjective sensation, mood, or mental task difficulty.

Which of 10 noxious stimuli is the more aversive? Which of 12 musical rhythms makes you feel more joyous? Which of 20 types of puzzle is the more difficult? In paired comparisons between each possible pair of stimuli, rhythms or puzzles, Hawkeye Harriet will pick what for her is the correct pair every time, so all we'll get from her is the rank-order of stimuli, rhythms and puzzles. Myopic Myra will less reliably and less accurately choose what for her is the correct pair, but her choice-rates will be correlated with how dissimilar each pair is. We'll recover much more precise information about the underlying structure of the stimulus set from error-prone Myra.

Annemarie's point about measurement is somewhat related to another fascinating phenomenon known as *stochastic resonance*. Briefly paraphrasing the Wikipedia entry for stochastic resonance (SR), SR occurs when a measurement or signal-detecting system's signal-to-noise ratio increases when a moderate amount of noise is added to the incoming signal or to the system itself. SR usually is observed either in bistable or sub-threshold systems. Too little noise results in the system being insufficiently sensitive to the signal; too much noise overwhelms the signal. Evidence for SR has been found in several species, including humans. For example, a 1996 paper in *Nature* reported a demonstration that subjects asked to detect a sub-threshold impulse via mechanical stimulation of a fingertip maximized the percentage of correct detections when the signal was mixed with a moderate level of noise. One way of thinking about the optimized version of Myopic Myra as a measurement instrument is to model her as a "noisy discriminator," with her error-rate induced by an optimal random noise-generator mixed with an otherwise error-free discriminating mechanism.