Modern psychophysics has traveled considerably beyond the threshold measures that dominated sensory studies in the first half of this century. Current methods capture the range of perceived intensity from threshold to maximum and promise to provide increasingly accurate comparisons of perceived intensities across individuals. The application of new psychophysical tools to genetic variation in taste allowed us to discover supertasters, individuals who live in particularly intense taste worlds. The percentage of supertasters varies with sex and with race. Females are more likely than are males to be supertasters. Asians are more likely than are Caucasians to be supertasters. Appropriate data on African Americans and Hispanics are not available. Because of the anatomy of the taste system, supertasters feel more bum from oral irritants like chili peppers, more creaminess/viscosity from fats and thickeners in food and may also experience more intense oral pain. In females, hormones further contribute to variation in taste. There is evidence that the ability to taste (particularly bitter) cycles with hormones in women of childbearing age, rises to a maximum early in pregnancy and declines after menopause. Not surprisingly, these sensory differences influence food choices and thus health. A discussion of the milestones on the road to understanding genetic variation in taste must include discussion of some potholes as well. Often our failures have been as instructive as our successes in the effort to evaluate the impact of genetic variation in taste. Our efforts to measure variation in taste experience have shed light on some problems with conventional psychophysical methods used to compare sensory (or hedonic) experiences. In order to compare these experiences across subjects (or groups), the experimenter must have a standard that can reasonably be assumed to be equal for all subjects (or on average for all groups). We use sensory standards (i.e., the method of magnitude matching). However, the "standards" used in many conventional scaling procedures are intensity adjectives (e.g., "weak," "medium," "strong"). Unfortunately, we now know that these adjectives do not indicate the same perceived intensities to all subjects. When data are treated as if the adjectives mean the same to all, differences across groups can be obscured or even reversed. This "reversal artifact" could contaminate a variety of studies.