aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

The Measurement of Pleasure and Pain

Perspectives on Psychological Science
2014, Vol 9(1) 91–93

DOI: 10.1177/1745691613512660
pps.sagepub.com
SAGE

Linda Bartoshuk
Department of Food Science and Human Nutrition, University of Florida

Abstract
Pleasure and pain are among our most salient experiences, and we want to know how our pleasures and pains stack up against those of others. Older psychophysical methods fail to provide valid comparisons of pleasure and pain across individuals or groups. We are making progress in measurement, but we still have a ways to go.

Keywords
pleasure, pain, individual differences, perception, psychophysics

William James wrote a letter to a colleague in 1898 in which he noted, "The whole subject of pleasure and pain and their dynamics is a very obscure one, in my opinion" (James, 1898/1971). At that time, there was already debate about the best way to measure pleasure and pain (e.g., see Major, 1895–1896) and that debate is still going on today. The deep problem is that, except for fictional mind readers, people cannot share each other's experiences of pleasure and pain. Yet with the misuse of scales, we sometimes act as if we can.

Traditional Scales Work for Comparing Within Subjects but Not Across Groups

The earliest category scale we know of dates back to the Greek astronomer Hipparchus: a 6-point scale to rate the brightness of stars. Various versions of category scales are still widely used to assess food preferences (1 = *dislike extremely*, 9 = *like extremely*) and pain (0 = *no pain*, 10 = *most intense pain*). Thanks to Aitken (1969), the category scale morphed into the visual analogue scale—often through the use of a line labeled in terms of the minimum and maximum intensities of a particular experience. The most common uses of both visual analogue scales and category scales is to make within-subject comparisons (e.g., Which coffee do you like best? Did your pain diminish after taking an analgesic?). This is a legitimate use.

However, these scales have also been used to make across-group comparisons (Do children like ice cream better than do adults? Do women experience more intense pain than do men?). This use is not legitimate and can lead to serious errors. Incidentally, this error keeps being rediscovered (see Aitken, 1969; Bartoshuk, Fast, & Snyder, 2005; Biernat & Manis, 1994; Birnbaum, 1999; Narens & Luce, 1983). Across-group comparisons are in error because they implicitly imply that we can share one another's experiences—we cannot. Your "extremely sweet" may be twice as sweet as "my extremely sweet." How can we find out?

Cross-Modality Matching

The answer evolved from the laboratory of S. S. Stevens at Harvard in the 1960s. The first step was cross-modality matching (Stevens, 1959). It turns out that we are very good at comparing perceived intensities across modalities (e.g., we can match the loudness of a tone to the brightness of a light).

I once heard psychophysicists described as people incapable of being bored. At the risk of boring readers kind enough to get this far, let me refer you to the references at the end (e.g., Bartoshuk et al., 2002) for some of the arcane discussion and proceed to the punch line. Once we know that we can compare sensory intensities across modalities, we have the core of a method to make across-group comparisons. If we can be sure that two sensory modalities are independent (i.e., variation in one is unrelated to variation in the other), then we can use

Corresponding Author:
Linda Bartoshuk, Department of Food Science and Human Nutrition, University of Florida, 4073 SW 21st Terrace, Gainesville, FL 32608
E-mail: lbartoshuk@dental.ufl.edu

one of them as a standard against which to look at variation in the other. Consider an example from taste. Tongues vary in the density of taste buds. Select two groups: those with high and low densities of taste buds. Ask both groups to rate the sweetness of a soda on a line labeled "no sweet" at one end and "extremely sweet" at the other end. Both groups rate the sweetness to be about two thirds of the way along the line. In the history of taste research, such a result has been used to argue that both groups experience the same sweetness from the soda.

Now we change to our newer method. We put earphones on everyone and ask both groups to adjust the loudness of a tone to the sweetness of the soda. If hearing and taste are really independent, then the average perception of loudness should be the same in the two groups, but subjects with the high density of taste buds set the tone to 90 decibels (about the loudness of a train whistle), and those with the low density set the tone to 80 decibels (about the loudness of a dial tone). As we know that a tone of 90 decibels sounds twice as loud as a tone of 80 decibels, we can conclude that the soda tastes twice as sweet to those with the high density of taste buds. We have identified a systematic difference in taste intensity between these two groups. Incidentally, individuals who experience the most intense taste sensations are called "supertasters." Of course, the key here is that our standard modality must be independent of the modality we wish to use for comparisons. In practice, it is wise to use multiple standard modalities.

Problems and Implications of Invalid Comparisons

What kinds of errors result from the invalid comparisons in earlier studies? For the most part, we fail to see differences that are real. For example, the new methods show that supertasters experience more pleasure from their favorite foods and more displeasure from their least favorite foods; the old 9-point category scale shows no such difference (Kalva, Sims, Puentes, Snyder, & Bartoshuk, in press). If your kids do not like vegetables, test their sense of taste—supertasters like vegetables less than do others (Dinehart, Hayes, Bartoshuk, Lanier, & Duffy, 2006).

Gilbert (2005) makes an argument with regard to happiness that has also been made by pain researchers:

> … [if] scales "are calibrated a bit differently for every person who uses them, then *it is impossible for scientists to compare the claims of two people*. That's a problem. But the problem isn't with the word *compare*, it's with the word two. Two is too small a number, and when it becomes two hundred or two thousand, the different calibrations of different individuals begin to cancel one another out. (p. 76)

Gilbert is right much of the time: Comparing experiences between two people is perilous, and increasing the sample size will help when there is no systematic difference between groups that affects the comparison. However, when there is a systematic difference, no increase in numbers will provide a valid comparison. The taste example shows that there are anatomical differences between supertasters and others that contribute to real differences in perception of taste intensity. No matter how many are tested, such systematic differences between supertasters and others will not "cancel each other out."

Consider another example: variable experiences across two groups. We asked several hundred individuals to rate perceived intensities of a variety of everyday experiences including "strongest pain of any kind experienced" (naming the source of the pain) and "brightest light ever seen" (Bartoshuk et al., 2004). For the women who selected childbirth as their worst pain, that pain was about 20% more intense than their brightest light. For men, their most intense pain was about equal to their brightest light. If we can assume no difference between women and men in the perception of brightness, then we have evidence that the strongest pain ever experienced is systematically different for these two groups.

Consider the clinical implications of our example. Let's say that a hospital sets the cut-off for providing analgesia at "4." For the women in our study who selected childbirth as their most intense pain, "10" denotes a more intense pain for them than for the men. Thus "4" would also denote a greater pain; these women would have to be in greater pain to get an analgesic in that hospital.

Some have tried to get around this by using "most intense pain imaginable" apparently assuming that we all share a common imaginary "most intense pain" no matter what we have actually experienced. Without the benefit of the fictional mind readers alluded to earlier, we cannot prove such an assumption, but simply asking subjects what their most intense imaginable sensation would be is illuminating. One suggested that being sucked into a black hole would qualify; some simply rated the most intense sensation they could imagine as equal to the most intense sensation they had experienced. Using "imaginable" adds noise to data but does not solve the problem of valid comparisons across individuals.

Conclusion

To the best of my knowledge, no one has gone through the literature tallying the number of errors in various fields resulting from invalid comparisons. I hope if

readers find such errors, they let me know. I would love to keep such a list.

To sum up, we can measure changes in pleasure and pain within an individual, but our old methods are not doing a good job of comparing differences in pleasure and pain across people. New methods to provide such comparisons are gaining ground (Coldwell et al., 2013). I hope someone chronicles that gain in another 25 years.

Declaration of Conflicting Interests

The author declared no conflicts of interest with respect to the authorship or the publication of this article.

References

Aitken, R. C. B. (1969). Measurement of feelings using visual analogue scales. *Proceedings of the Royal Society of Medicine*, *62*, 989–993.

Bartoshuk, L. M., Duffy, V. B., Chapo, A. K., Fast, K., Yiee, J. H., Hoffman, H. J., . . . Snyder, D. J. (2004). From psychophysics to the clinic: Missteps and advances. *Food Quality and Preference*, *15*, 617–632.

Bartoshuk, L. M., Duffy, V. B., Fast, K., Green, B. G., Prutkin, J. M., & Snyder, D. J. (2002). Labeled scales (e.g., category, Likert, VAS) and invalid across-group comparisons: What we have learned from genetic variation in taste. *Food Quality and Preference*, *14*, 125–138.

Bartoshuk, L. M., Fast, K., & Snyder, D. (2005). Differences in our sensory worlds: Invalid comparisons with labeled scales. *Current Directions in Psychological Science*, *14*, 122–125.

Biernat, M., & Manis, M. (1994). Shifting standards and stereotype-based judgments. *Journal of Personality and Social Psychology*, *66*, 5–20.

Birnbaum, M. H. (1999). How to show that 9 > 221: Collect judgments in a between-subjects design. *Psychological Methods*, *4*, 243–249.

Coldwell, S. E., Mennella, J. A., Duffy, V. B., Pelchat, M. L., Griffith, J. W., Smutzer, G., . . . Hoffmann, H. J. (2013). Gustation assessment using the NIH Toolbox. *American Academy of Neurology*, *80*, S20–S24.

Dinehart, M. E., Hayes, J. E., Bartoshuk, L. M., Lanier, S. L., & Duffy, V. B. (2006). Bitter taste markers explain variability in vegetable sweetness, bitterness, and intake. *Physiology & Behavior*, *87*, 304–313.

Gilbert, D. (2005). *Stumbling on happiness*. New York, NY: Random House.

James, W. (1971). *A letter on pleasure and pain*. Los Angeles, CA: W. Nethery. Original work published 1898.

Kalva, J. J., Sims, C. A., Puentes, L. A., Snyder, D. J., & Bartoshuk, L. M. (in press). Comparison of the hedonic general Labeled Magnitude scale with the hedonic 9-point scale. *Journal of Food Science*.

Major, E. R. (1895–1896). On the affective tone of simple sense-impressions. *American Journal of Psychology*, *7*, 57–77.

Narens, L., & Luce, R. D. (1983). How we may have been misled into believing in the interpersonal comparability of utility. *Theory and Decision*, *15*, 247–260.

Stevens, J. C. (1959). Cross-modality validation of subjective scales for loudness, vibration, and electric shock. *Journal of Experimental Psychology*, *57*, 201–209.