More than just a name:

The psychological effects of ‘Statistical Significance’

V.N. Vimal Rao

University of Minnesota

# Abstract

*Traditional statistical hypothesis testing privileges p-values less than .05 as achieving so-called statistical significance. Since the mid-20th century, this practice has drawn heavy theoretical and methodological criticism, culminating in a 2019 declaration by the American Statistical Association that the term statistical significance should henceforth be abandoned. Lost in theoretical and methodological debates is the psychological consequences of the statistical significance category label on individuals’ mental representations of p-values. Category labels can incite a distortion of the mental representation of physical continuums, known as a categorical perception effect, where differences within a category are perceptually minimized while differences across categorical boundaries are perceptually exaggerated. We examined whether .05 has become a such a boundary, forming a discontinuity in the mental number line, exaggerating perceived differences between statistically significant p-values and those that are not. 25 graduate students with statistical training viewed pairs of p-values and judged whether they were “similar” or “different”. Participants were between 1.5 to 4 times more likely to judge p-values as “different” when they crossed the .05 boundary (e.g., .048* – *.052) compared to when they did not (e.g., .028* – *.032). Furthermore, participants were approximately 24% faster at selecting “different” when p-values crossed the .05 boundary. This evidence, consistent with the predictions of a categorical perception effect and contrary to current theories of the mental representation of numbers, suggests that traditional statistical training creates a psychologically real separation between statistically significant and non-significant p-values. A distortion of the mental representation of p-values implies that statistical significance, so entrenched in teaching and practice, is more than just a name – it has permeated the way we fundamentally perceive p-values, a precursor to reasoning about them and making inferences. To truly realize a world beyond p < .05, we must first rectify the psychological repercussions of a century of statistical significance.*

Keywords: statistics education; rational number processing; categorical perception.

# References

Cohen, D. J. (2010). Evidence for direct retrieval of relative quantity information in a quantity judgment task: Decimals, integers, and the role of physical similarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(6), 1389.

Cudeck, R. (1996). Mixed-effects models in the study of individual differences with repeated measures data. *Multivariate behavioral research, 31*(3), 371-403.

Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *Journal of experimental Psychology: Human Perception and performance, 16*(3), 626.

Fisher, R.A. (1925). *Statistical methods for research workers.* London: Oliver and Boyd.

Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception & psychophysics, 66*(3), 363-376.

Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General, 123*(2), 178.

Harnad, S. (1987) Psychophysical and cognitive aspects of categorical perception: A critical overview. In Harnad, S. (ed.) (1987) *Categorical Perception: The Groundwork of Cognition*. New York: Cambridge University Press.

Harnad, S. (2017). To cognize is to categorize: Cognition is categorization. In *Handbook of categorization in cognitive science* (pp. 21-54). Elsevier.

Hary, J. M., & Massaro, D. W. (1982). Categorical results do not imply categorical perception. *Perception & Psychophysics, 32*(5), 409-418.

Krueger, J. I., & Heck, P. R. (2019). Putting the p-value in its place. *The American Statistician, 73*(S1), 122-128.

LeFevre, J. A., Sadesky, G. S., & Bisanz, J. (1996). Selection of procedures in mental addition: Reassessing the problem size effect in adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(1), 216.

Marschner, I. C., & Gillett, A. C. (2012). Relative risk regression: reliable and flexible methods for log-binomial models. *Biostatistics, 13*(1), 179-192.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language, 94,* 305-315.

Notman, L. A., Sowden, P. T., & Özgen, E. (2005). The nature of learned categorical perception effects: a psychophysical approach. *Cognition, 95*(2), B1-B14.

Nuerk, H. C., Moeller, K., Klein, E., Willmes, K., & Fischer, M. H. (2011). Extending the mental number line: A review of multi-digit number processing. *Journal of Psychology, 219*(1), 3.

Nuerk, H. C., Weger, U., & Willmes, K. (2001). Decade breaks in the mental number line? Putting the tens and units back in different bins. *Cognition, 82*(1), B25-B33.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Schulze, K. G., Schmidt-Nielsen, A., & Achille, L. B. (1991). Comparing three numbers: The effect of number of digits, range, and leading zeros. *Bulletin of the Psychonomic Society, 29*(4), 361-364.

Varma, S., & Karl, S. R. (2013). Understanding decimal proportions: Discrete representations, parallel access, and privileged processing of zero. *Cognitive Psychology, 66*(3), 283-301.

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a World Beyond “p< 0.05”. *The American Statistician, 73* (S1), 1-19.