Mental representation of *p*-values:

Does a categorical perception effect exist?

Yes.

"*p* < 0.05" is a convenient rule, but it warps our perception of *p*-values

V.N. Vimal Rao, Jeffrey K. Bye, Sashank Varma

COLLEGE OF EDUCATION + HUMAN DEVELOPMENT

Department of Educational Psychology

UNIVERSITY OF MINNESOTA
Driven to Discover®

Background

The world of "*p* < 0.05"

- The *p*-value is a quantitative tool to measure the falsifiability of a null hypothesis.
- Students are often taught to reject the null hypothesis when "*p* < 0.05", a level popularized by R.A. Fisher in 1925 as a convenient threshold.
- Such categories can facilitate learning and extraction of information, but they also can lead to distortions of the mental representation of the *p*-value continuum, called a categorical perception effect.

Categorical perception effects

- A categorical perception effect (CPE) is a difference in how a set of stimuli are perceived depending on if they are classified as belonging to the same category.
- A CPE manifests as a boundary effect and within-category indiscriminability.
- A **boundary effect** means that *p*-values less than 0.05 receive one label while *p*-values greater than 0.05 receive a different label.
- **Within-category indiscriminability** means that individuals are less likely to call a pair of *p*-values different if they are both less than (or both greater than) 0.05 compared to if they are on different sides.

The study design

- 40 participants were recruited via a convenience sample, of which an eligible 25 completed the study.
- All participants had completed at least one year of coursework in statistics at the graduate level.
- Pairs of *p*-values were created **controlling for distance**, **size**, & **compatibility effects** by condition: across 0.05, both below 0.05, and both above 0.05.
- Participants were shown a *p*-value for 1 sec, then were shown a second *p*-value and asked to indicate whether they were similar or different by pressing either 'F' or 'J' on their keyboards.
- The next stimuli pair within the block was then shown after a brief delay.
- 108 pairs of *p*-values were shown in six blocks, with a 30 second break between each block. The study took approximately 10 minutes to complete in full.
- Participants' selections and reaction time were collected via Qualtrics.

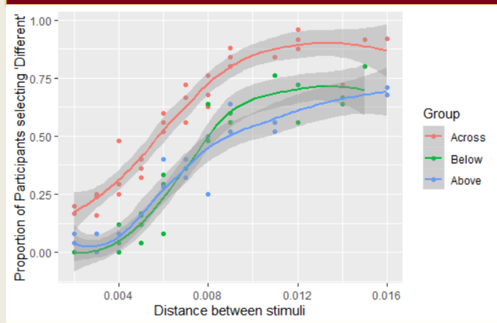
Results and Implications

- **Individuals are more likely to indicate that a pair of *p*-values are different when they cross the 0.05 boundary and are slightly faster at labelling a pair as different when they do cross the boundary. This evidence is consistent with a CPE hypothesis.**
- This study's results, providing a first insight into the cognitive models underlying interpretations of *p*-values, set the stage for future research in education and training needed for students to benefit from the dichotomization of the p-value continuum while ensuring that they do not fall into the traps of dichotomization and categorical perception effects.
- The CPE is an important consideration to make when evaluating and discussing the *p*-value controversy and phenomena such as *p*-hacking.

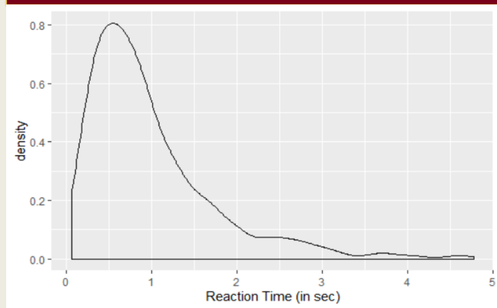
Future directions

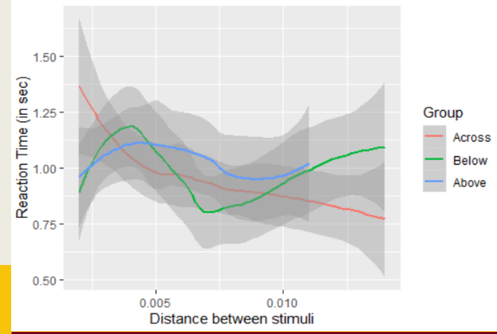
- This study used an AX discrimination task, whereby participants compared a p-value to a single referent p-value. We can also use an ABX discrimination task, in which participants are shown one p-value below 0.05 and one above 0.05 and asked to determine which of those two a third p-value is most like.
- It is likely that the mental representations of p-values of novices and experts differ. We can explore these individual differences in the CPEs.
- CPEs can be learned, as well as unlearned. We can assess alternate models and sequences of instruction for p-values in terms of students' CPEs.

Discriminability



Reaction Time





Statistical Models

Log-Binomial Mixed-Effects model

- **Participants were approximately 36.6% more likely to indicate a pair of *p*-values were *different* when the values were on different sides of the 0.05 boundary (*p* = 0.013; 95% CI: 6.9% to 74.7%).**

Linear Mixed-Effects model for RT

- **Participants were approximately 0.07 secs faster at identifying pairs of stimuli as different when they crossed the boundary (*se* = 0.04 secs; *p* = 0.078).**
- **This is approximately a 6% reduction RT against the model adjusted mean RT of 1.16 seconds.**

Additional Materials