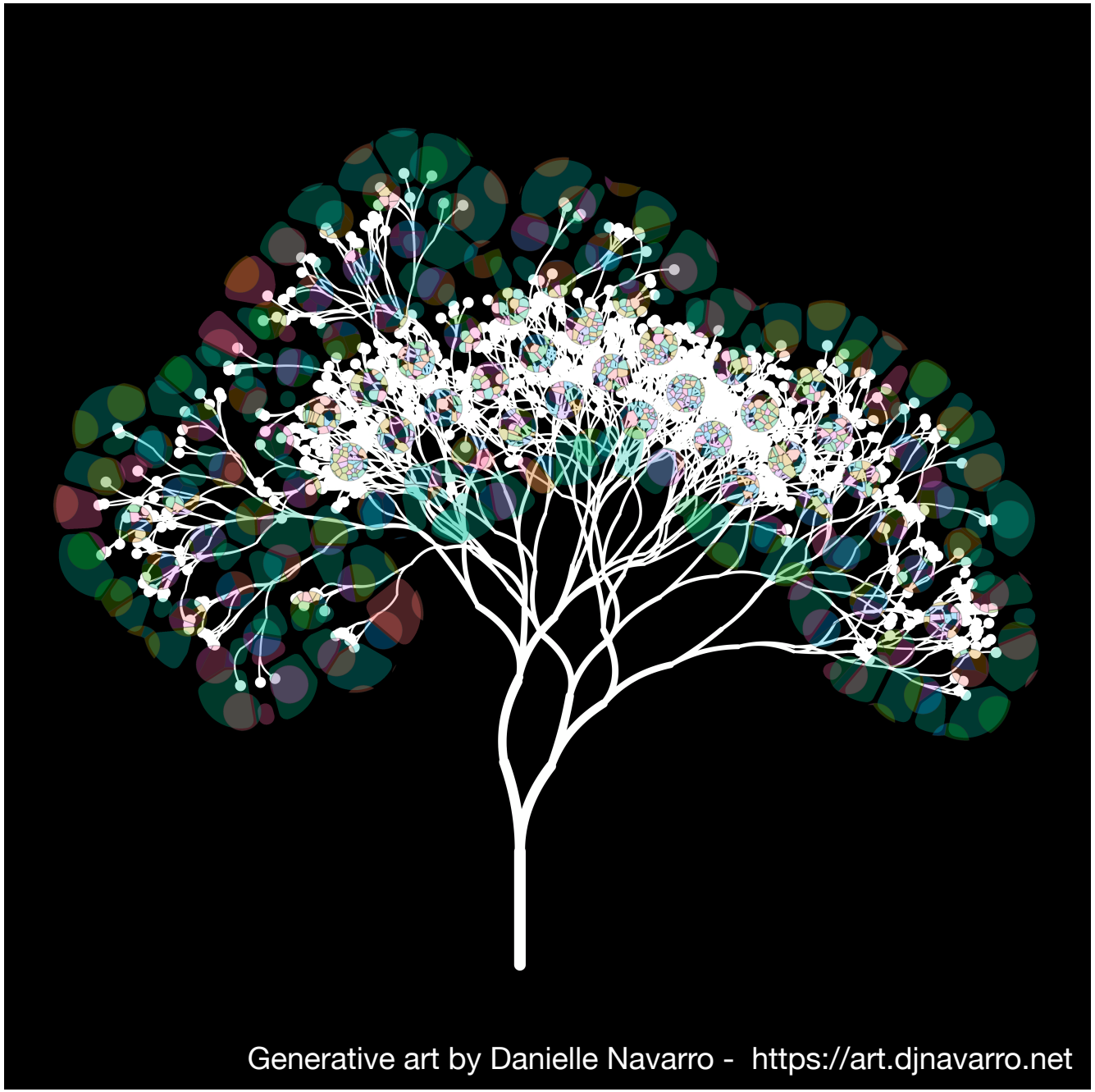
Power for detecting the presence of set size differences in the contralateral delay activity

William Ngiam, Kirsten Adam, Colin Quirk, Edward Vogel, Edward Awh

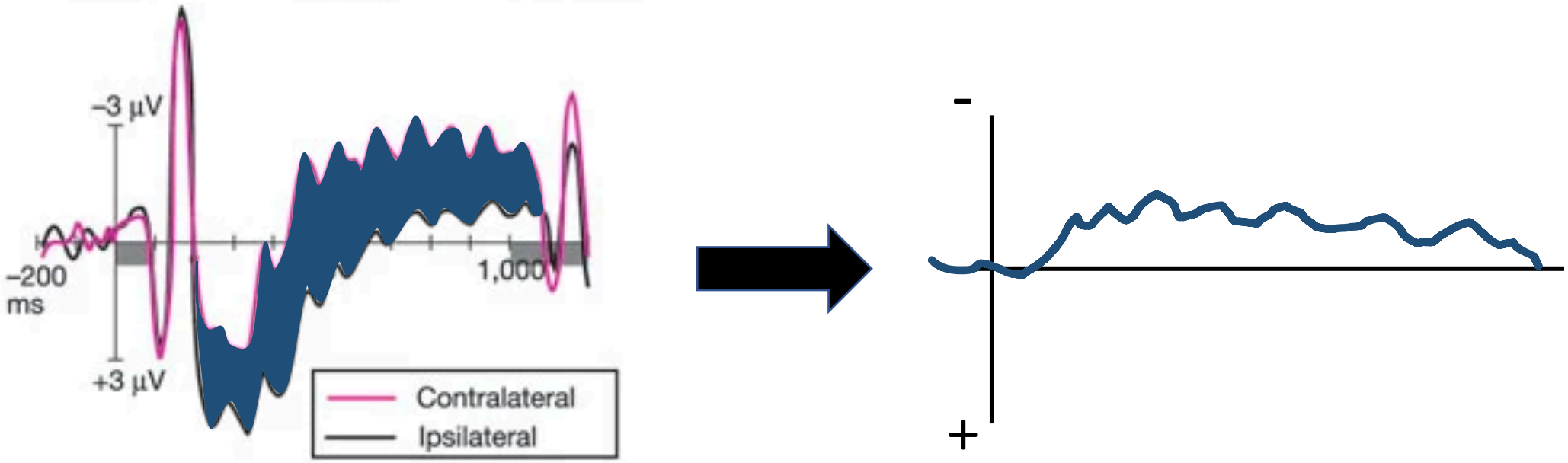



THE UNIVERSITY OF CHICAGO



Contralateral Delay Activity (CDA)

- An event-related potential (ERP) measured during the retention period of a bilateral change-detection task
  - Sustained negativity over the contralateral hemisphere
  - Across posterior and occipital electrodes
  - Mean amplitude of difference wave between contralateral and ipsilateral electrodes

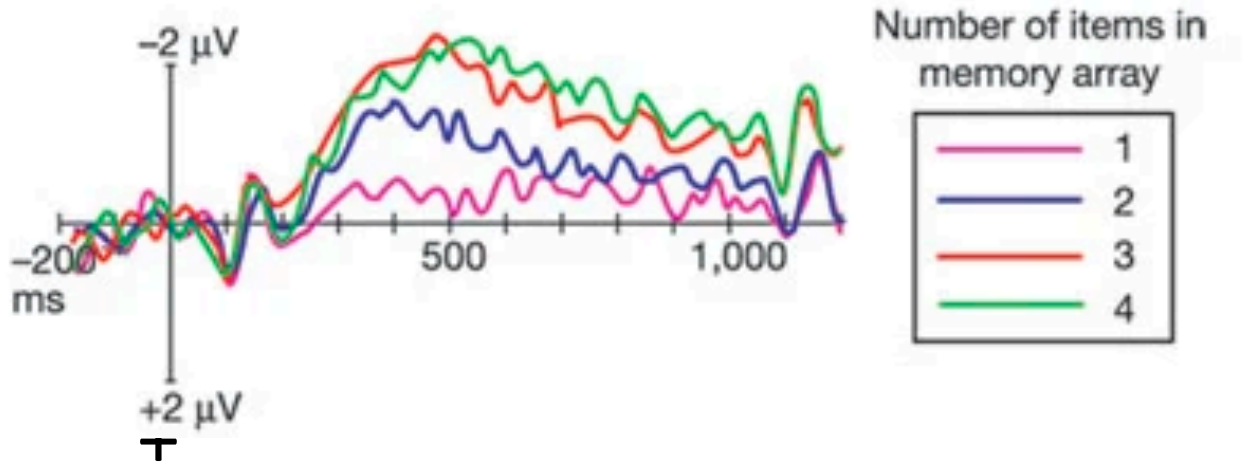


@will_ngiam; wngiam@uchicago.edu

Vogel & Machizawa (2004). *Nature*, *428*(6984).

Contralateral Delay Activity (CDA)

- Closely tracks the memory load
  - Increases in amplitude with increasing set-size, plateauing at a typical VWM capacity of 3-4 items



@will_ngiam; wngiam@uchicago.edu

Vogel & Machizawa (2004). *Nature*, *428*(6984).

Contralateral Delay Activity (CDA)

- Useful tool in advancing research on visual working memory
  - Filtering efficiency and attention
  - Binding and grouping
  - Complexity and resolution
  - Visual search
  - Multiple object tracking
  - Rapid-serial visual presentation

- For a review: Luria, Balaban, Awh, & Vogel (2016). *Neuroscience & Biobehavioral Reviews*, 62, 100-108.

@will_ngiam; wngiam@uchicago.edu

What is power?

- The probability of detecting an effect given the effect exists
  - *p*(significant|effect)

- As scientists, we want to maximize the positive predictive value - the probability that a positive (significant) finding is due to a true effect and not a false positive
  - *p*(effect|significant)

- The positive predictive value is maximized by increasing our statistical power
  - A potential reproducibility crisis in science due to a high prevalence of false positives

@will_ngiam; wngiam@uchicago.edu

Ioannidis (2005). *PLos med*, *2*(8), e124.
Open Science Collaboration. (2015). *Science*, 349(6251), aac4716.

Why we need to care about power

- It's important that experimenters are sufficiently powered to detect the effect of interest
  - Researchers consume precious time and resources in conducting the experiment
  - A false negative (due to lack of power) can mean:
    - A slowing of scientific progress as the negative results cannot be reliably interpreted
    - Left in a file drawer as the result is a null effect
    - And we have missed out on knowing the effect exists!

- Increasing statistical power improves the reliability of the findings
  - Estimated median statistical power is between 8 and 31% due to typically small sample sizes in neuroscience (Button et al., 2013)

- Estimating the number of subjects and trials needed is useful!
  - Enables pre-registration of sample sizes and study design
  - Evaluate whether the experiment is feasible

@will_ngiam; wngiam@uchicago.edu

Ioannidis. (2005). *PLos med*, *2*(8), e124.
Button, Ioannidis, Mokrysz, Nosek, Flint, Robinson, & Munafò (2013). *Nature Reviews Neuroscience*, 14(5), 365-376.

Estimating power to measure the CDA

- What number of subjects and number of trials is required to be sufficiently powered to detect:
  - The presence/absence of the CDA
  - Set-size differences in the CDA

@will_ngiam; wngiam@uchicago.edu

Datasets

- Unsworth et al. (2015) *JoCN* dataset
  - Color working memory task with set size 2 and 6
  - 171 participants – very large EEG dataset
  - 139 participants had at least 170 clean trials in both set size conditions

- Hakim et al. (2019) *PsychSci* dataset
  - Color working memory task with set size 2 and 4
  - 31 participants with at least 220 clean trials in both set size conditions

@will_ngiam; wngiam@uchicago.edu

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2015). *Journal of Cognitive Neuroscience*, 27(5), 853-865.
Hakim, N., Adam, K. C., Gunseli, E., Awh, E., & Vogel, E. K. (2019). *Psychological science*, 30(4), 526-540.

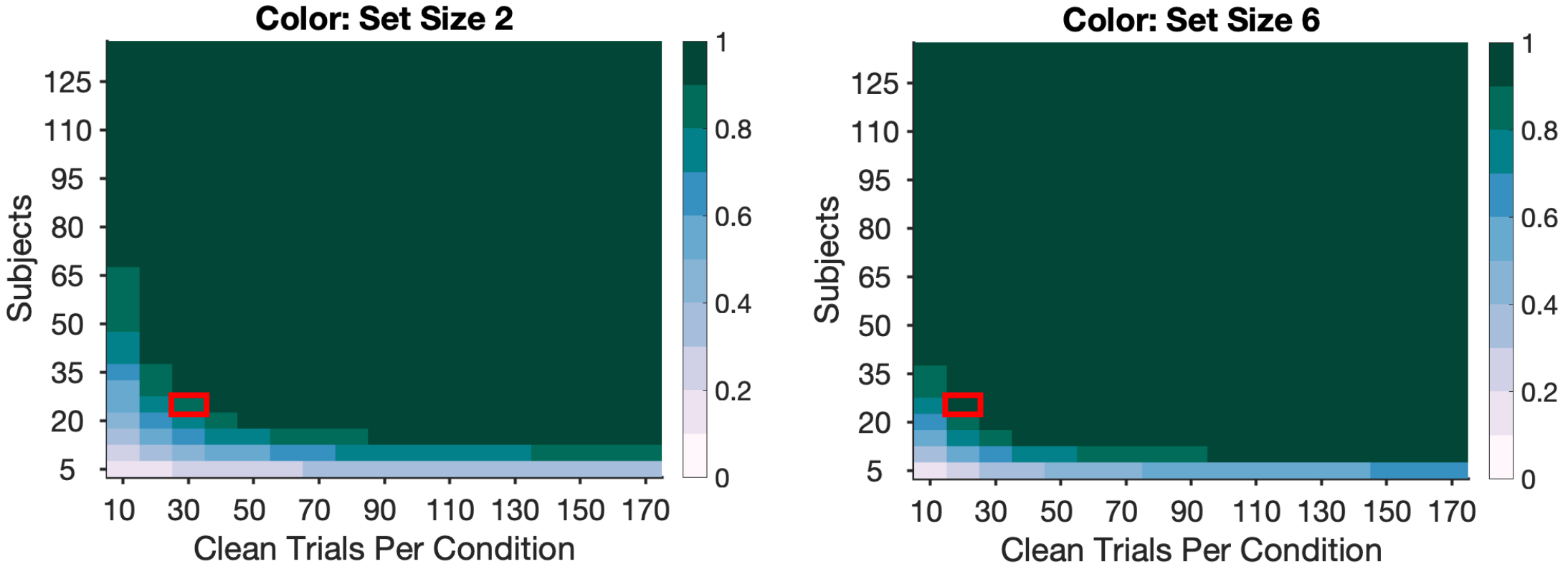
Subsampling analysis

- Randomly sampled a given number of trials from a given number of randomly sampled subjects
  - Conducted a statistical test on the subsampled data
  - 10,000 iterations of each combination of number of trials and subjects
  - Estimate power with the proportion of iterations that produced significant tests

@will_ngiam; wngiam@uchicago.edu

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2015). *Journal of Cognitive Neuroscience*, 27(5), 853-865.

Power to detect the presence of CDA

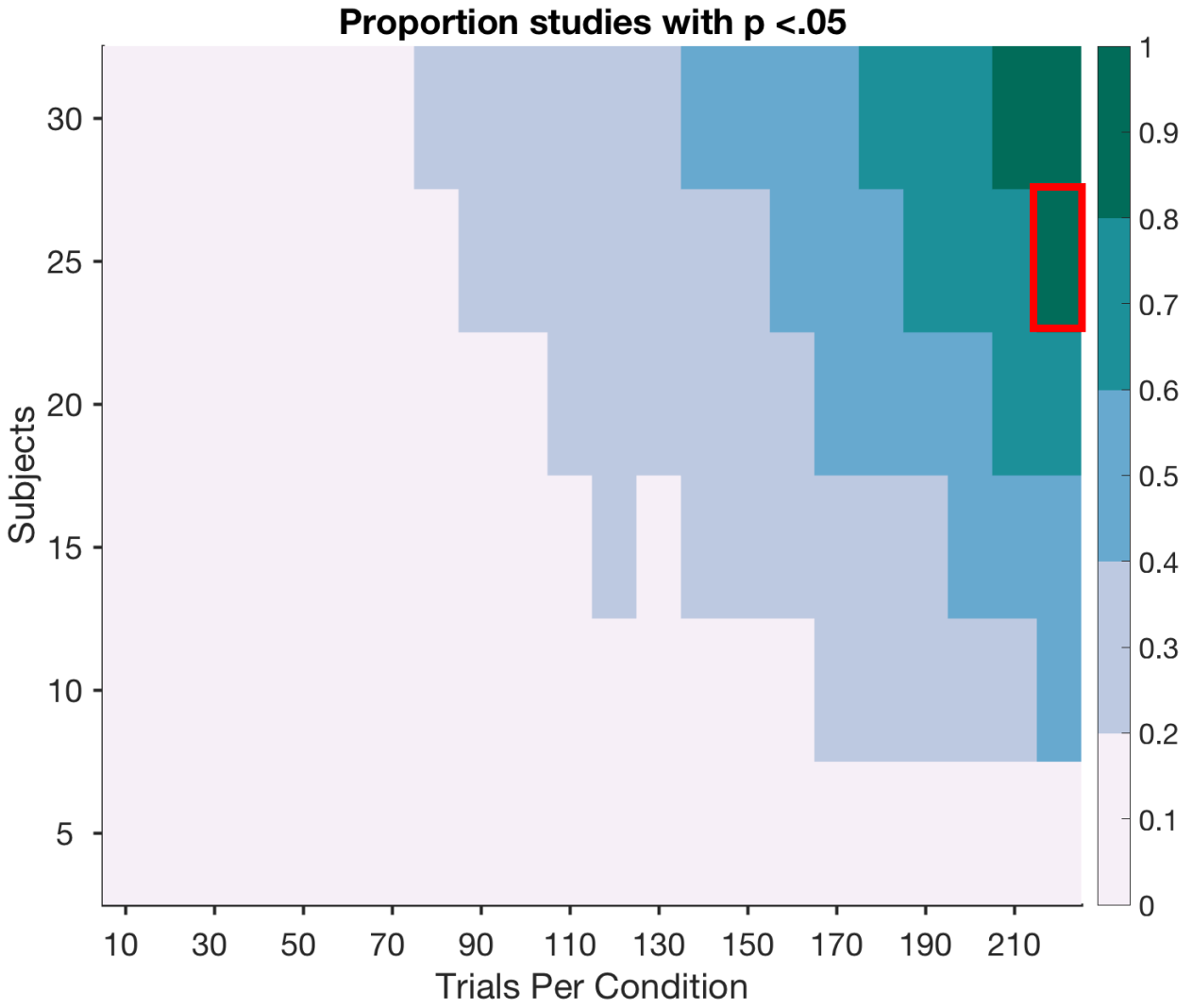


- For $n = 25$
  - 30 trials required for at least 80% power in set size 2
  - 20 trials required for at least 80% power in set size 6

@will_ngiam; wngiam@uchicago.edu

Power to detect set-size differences in CDA

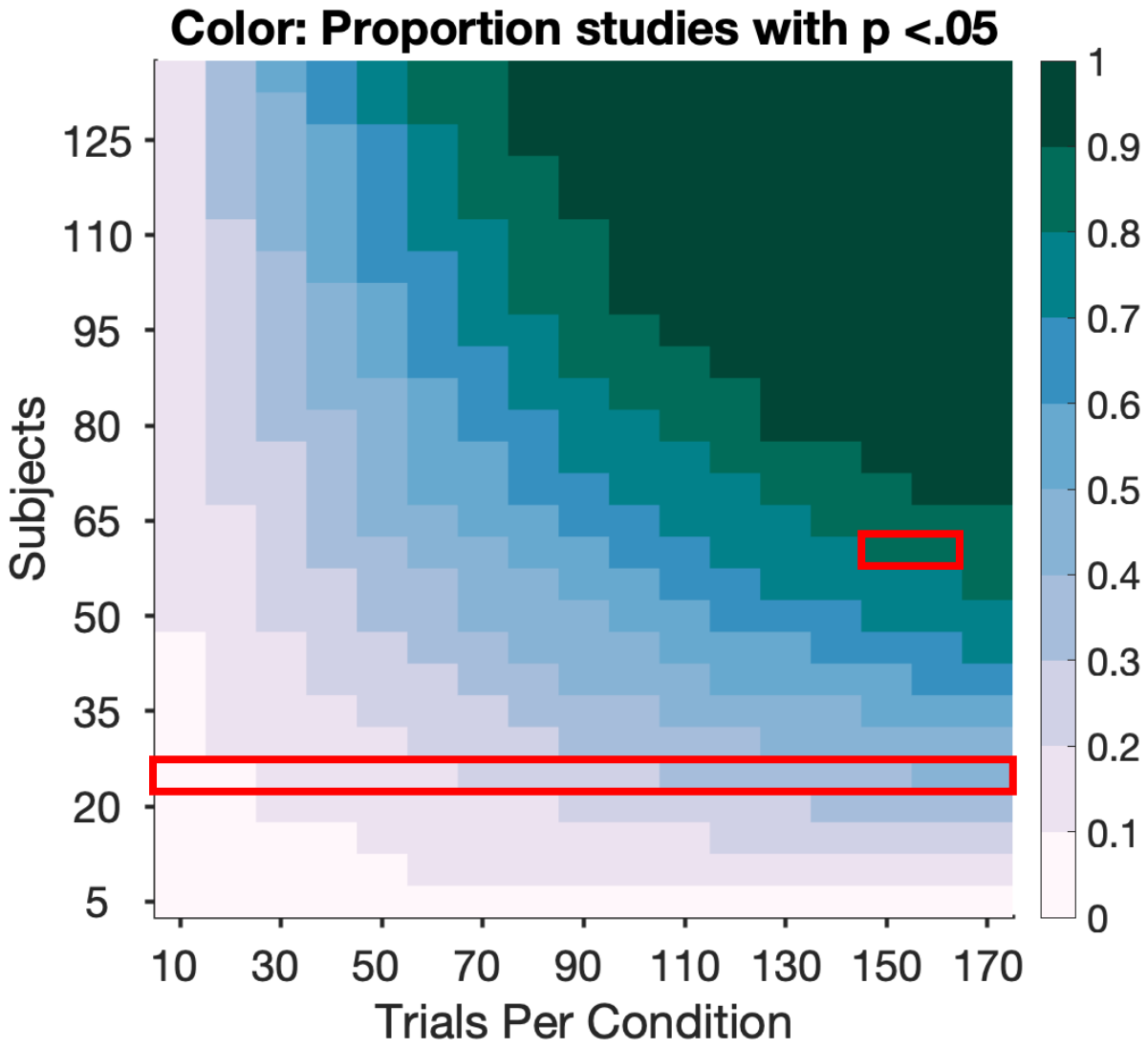
- To detect a 2 vs 4 set-size difference in the CDA:
  - For *n* = 25, about 220 clean trials is needed to achieve 80% power



@will_ngiam; wngiam@uchicago.edu

Power to detect set-size differences in CDA

- To detect a 2 vs 6 set-size difference in the CDA:
  - For *n* = 25, 80% power is not achieved with 170 clean trials
  - To achieve 80% power, you might need 60 subjects and 150 clean trials

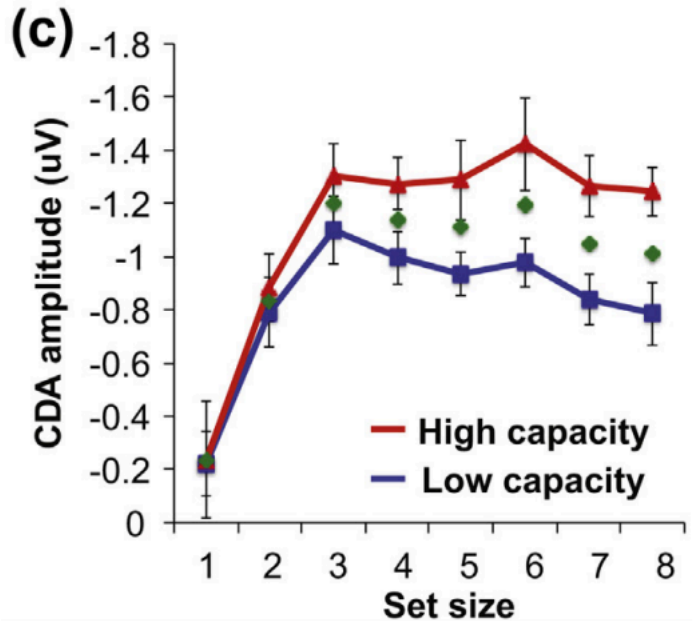


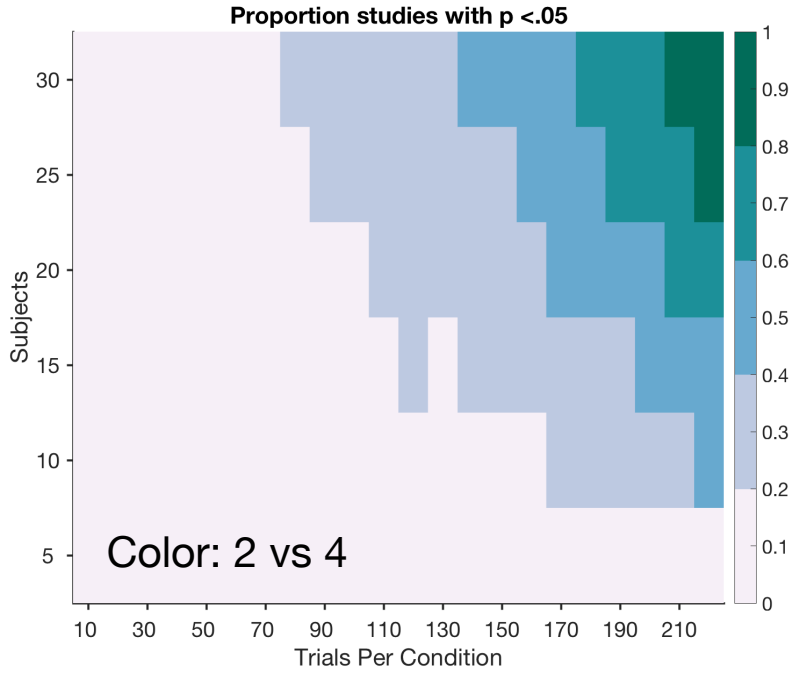
@will_ngiam; wngiam@uchicago.edu

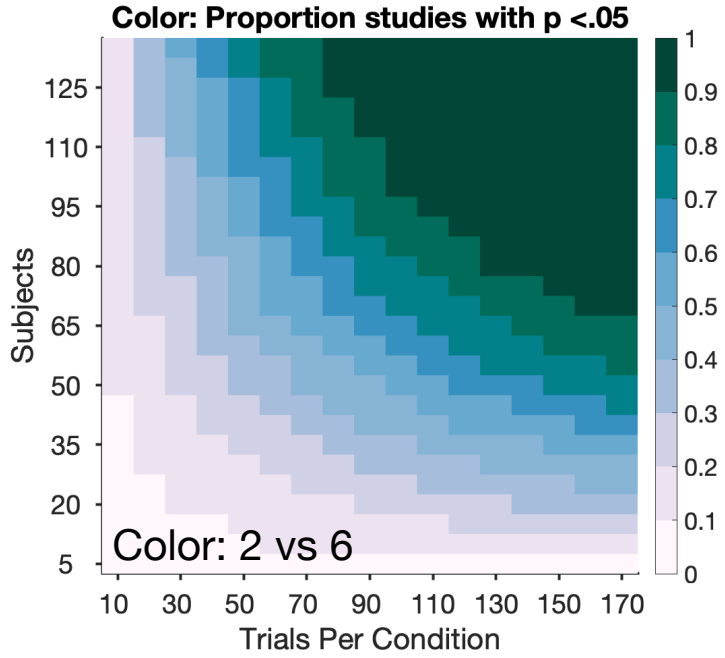
Power to detect set-size differences in CDA

- Set-size effects between 2 versus 6 appear to be harder to detect compared to 2 versus 4 (smaller effect size)
  - Set size 6 may overload the VWM system and recruit filtering processes

- Suggests experimental designs that target 1 versus 3 instead for example







@will_ngiam; wngiam@uchicago.edu

Fukuda, Woodman, & Vogel (2015). *Mechanisms of sensory working memory: attention and performance XXV.*

Limitations

- These analyses are restricted to a maximum of 200 clean trials
  - We have preliminary simulations predicting beyond these limits
- How well these analyses extrapolate relies on the generalizability of these datasets
  - Given these are larger than standard EEG datasets suggests it has more external validity
- Power can be improved with more precise measurement of the CDA (see Luck et al., 2020 for measure of ERP data quality)

Takeaways

- Typical numbers of subjects and trials for contralateral delay activity experiments may be underpowered when detecting set size differences

- The size of the effect is critical when planning numbers of subjects and trials in any experiment
  - Improve your experimental design to increase power

- We hope this subsampling analysis will enable principled estimates of power, ensuring better chances of detecting the effect
  - We have preliminary simulations beyond the number of subjects and trials

@will_ngiam; wngiam@uchicago.edu

Thank you!

- I would love to hear your feedback
  - Twitter: @will_ngiam
  - Email: wngiam@uchicago.edu

UC San Diego

Kirsten Adam

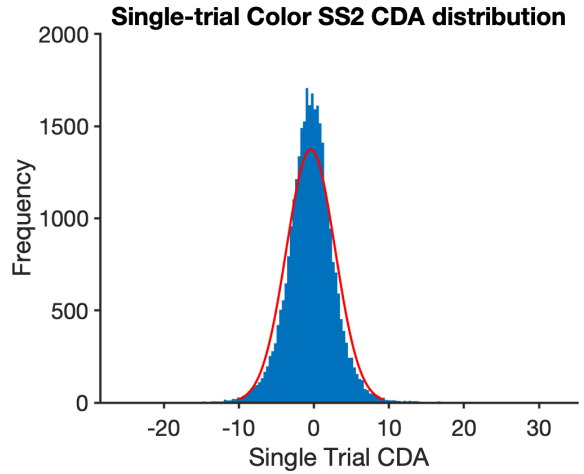
THE UNIVERSITY OF CHICAGO

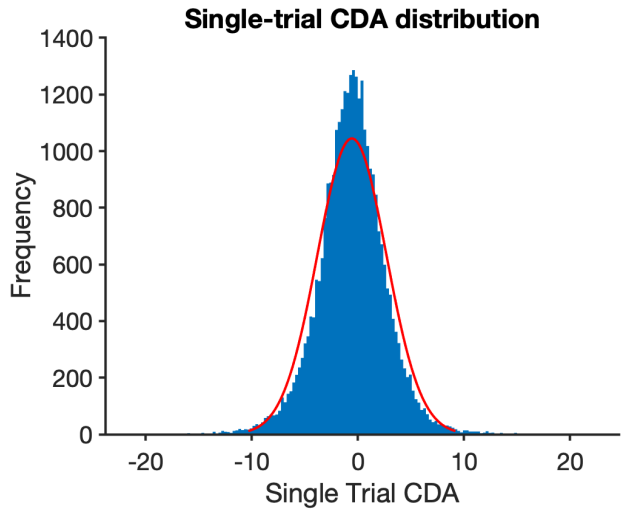
Colin Quirk

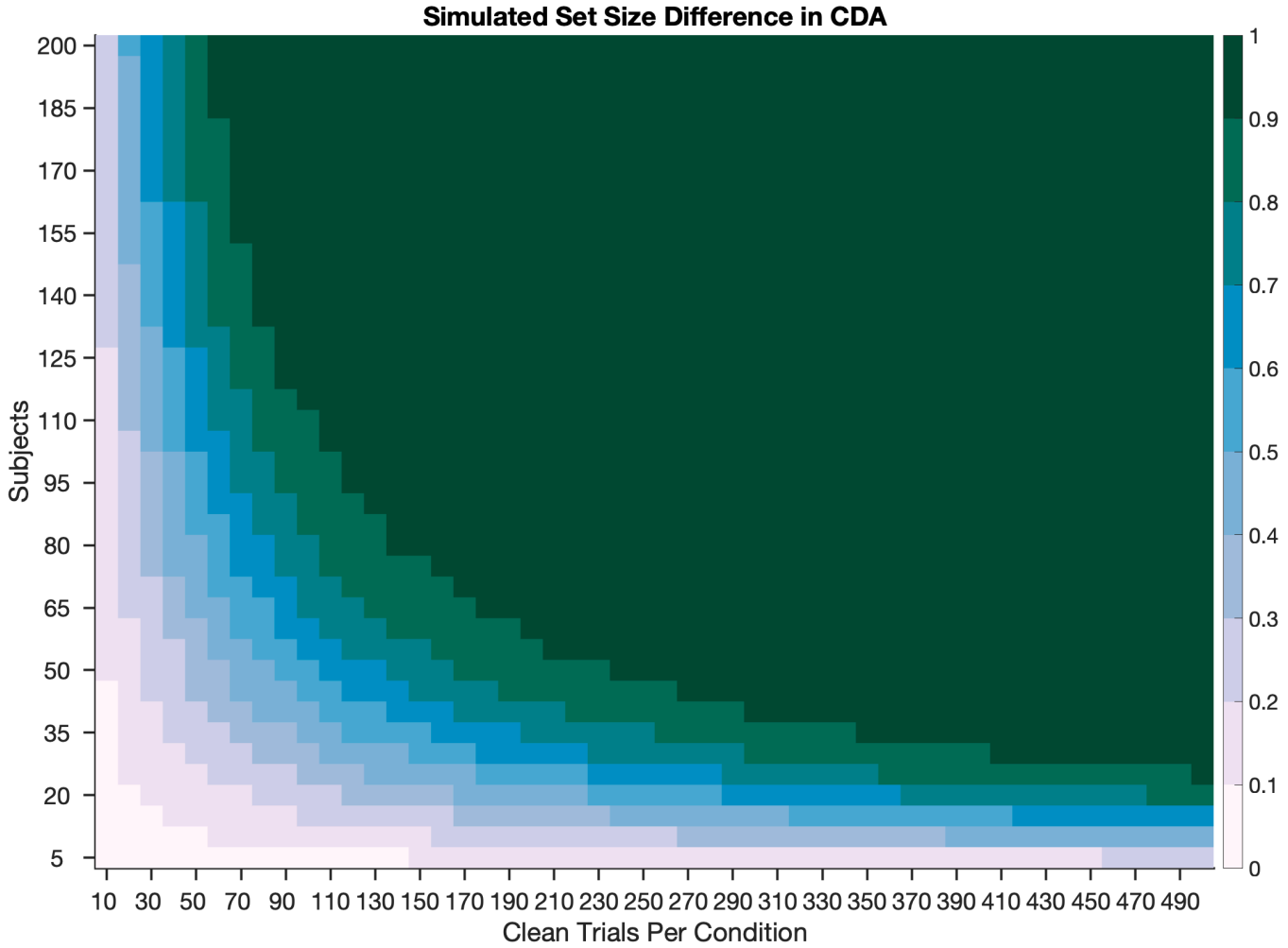
Edward Vogel

Edward Awh

Simulations







@will_ngiam; wngiam@uchicago.edu