Title

Estimating the statistical power to detect set-size effects in contralateral delay activity

Journal

Psychophysiology, 58(5)

ISSN

0048-5772

Authors

Ngiam, William XQ
Adam, Kirsten CS
Quirk, Colin
et al.

Publication Date

2021-05-01

DOI

10.1111/psyp.13791

Peer reviewed

HHS Public Access

Author manuscript

Psychophysiology. Author manuscript; available in PMC 2021 May 01.

Published in final edited form as:

Psychophysiology. 2021 May ; 58(5): e13791. doi:10.1111/psyp.13791.

Estimating the statistical power to detect set-size effects in contralateral delay activity

William X. Q. Ngiam[1], **Kirsten C. S. Adam**[2], **Colin Quirk**[1], **Edward K. Vogel**[1], **Edward Awh**[1]

[1]Department of Psychology, University of Chicago, Chicago, IL, USA

[2]Department of Psychology, University of California San Diego, San Diego, CA, USA

Abstract

The contralateral delay activity (CDA) is an event-related potential component commonly used to examine the online processes of visual working memory. Here, we provide a robust analysis of the statistical power that is needed to achieve reliable and reproducible results with the CDA. Using two very large EEG datasets that examined the contrast between CDA amplitude with set sizes 2 and 6 items and set sizes 2 and 4 items, we present a subsampling analysis that estimates the statistical power achieved with varying numbers of subjects and trials based on the proportion of significant tests in 10,000 iterations. We also generated simulated data using Bayesian multilevel modeling to estimate power beyond the bounds of the original datasets. The number of trials and subjects required depends critically on the effect size. Detecting the *presence* of the CDA—a reliable difference between contralateral and ipsilateral electrodes during the memory period—required only 30-50 clean trials with a sample of 25 subjects to achieve approximately 80% statistical power. However, for detecting a difference in CDA amplitude between two set sizes, a substantially larger number of trials and subjects were required; approximately 400 clean trials with 25 subjects to achieve 80% power. Thus, to achieve robust tests of how CDA activity differs across conditions, it is essential to be mindful of the estimated effect size. We recommend researchers designing experiments to detect set-size differences in the CDA collect substantially more trials per subject.

Keywords

contralateral delay activity; EEG; ERPs; statistical power; visual working memory

Correspondence: William X. Q. Ngiam, Department of Psychology, University of Chicago, 940 E 57th Street, Chicago, IL, USA. wngiam@uchicago.edu.

AUTHOR CONTRIBUTIONS

William X. Q. Ngiam: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Software; Visualization; Writing-original draft; Writing-review & editing. **Kirsten C. S. Adam:** Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Software; Visualization; Writing-review & editing. **Colin Quirk:** Formal analysis; Investigation; Methodology; Software; Visualization. **Edward K. Vogel:** Conceptualization; Funding acquisition; Methodology; Supervision. **Edward Awh:** Conceptualization; Funding acquisition; Methodology; Project administration; Supervision; Writing-original draft; Writing-review & editing.

1 | INTRODUCTION

The retention of information for ongoing visual cognition is known as visual working memory (VWM). Although our visual perception appears to be extremely rich, the capacity for VWM is surprisingly limited to approximately three items' worth (Adam et al., 2017; Cowan, 2001; Luck & Vogel, 1997). With strong associations to measures of fluid intelligence (Fukuda et al., 2010; Unsworth et al., 2015) and overall cognition (Luck & Vogel, 2013), there has been continued interest from researchers in understanding the factors that influence VWM capacity.

A popular tool for examining capacity limits in VWM is the contralateral delay activity (CDA), an event-related potential (ERP) component measured using electroencephalography (EEG). The CDA is a sustained negative wave that emerges across the posterior and occipital electrodes during the retention period of a VWM task (Luria et al., 2016; Vogel & Machizawa, 2004). This wave tracks closely with VWM storage (Feldmann-Wüstefeld et al., 2018; Hakim et al., 2019), increasing in amplitude with the set size of the memory display until it plateaus at the VWM capacity limit of approximately three to four items (Fukuda et al., 2015; Ikkai et al., 2010; Vogel & Machizawa, 2004). This memory load signal has proven to be useful as a window into online VWM processes that cannot be easily measured with a behavioral response at the end of the trial. For example, the CDA has been used to examine the influence of attention and long-term memory on working memory maintenance (Asp et al., 2019; Brady et al., 2016; Göddertz et al., 2018; Heuer & Schubö, 2016; Kuo et al., 2012; Quirk et al., 2020; Salahub et al., 2019; Williams & Woodman, 2012; Xie & Zhang, 2017) and to demonstrate the role of VWM in guiding visual search (Carlisle et al., 2011; Emrich et al., 2009; Emrich et al., 2010; Gunseli et al., 2014; Williams & Drew, 2020; Woodman & Arita, 2011), in mental rotation (Prime & Jolicoeur, 2009), in attentional control (Vogel et al., 2005), in multiple object tracking (Drew et al., 2011), and in grouping (Balaban & Luria, 2015; Rabbitt et al., 2017). Likewise, the CDA has been used to track changes to working memory performance in clinical populations (Lee et al., 2010; Leonard et al., 2013; Meconi et al., 2014; Qi et al., 2014; Salahub & Emrich, 2019; Spronk et al., 2013; Stout et al., 2013; Wiegand et al., 2016; Zaehle et al., 2013) and across the lifespan (Astle et al., 2014; Duarte et al., 2013; Jost et al., 2011; Pagano et al., 2015; Sander et al., 2011; Spronk et al., 2012). In sum, load-dependent changes in CDA amplitude have been widely replicated, and the CDA has been fruitfully used to address a wide range of basic and applied questions. Nevertheless, in this article, we turn a critical lens on the potential methodological issue which may impede our ability to observe reliable effects with the CDA.

The effective use of the CDA, however, depends critically on collecting sufficient quantities of data to ensure robust estimates of the component. More broadly, there is a notable cause for concern that the majority of the published scientific literature may be false, due to the lack of statistical power leading to a high prevalence of false-positive findings (Ioannidis, 2005). Estimates of statistical power have been worryingly low—approximately 24% across science in the past 60 years (Smaldino & McElreath, 2016) and between 8% and 31% across neuroscience disciplines (Button et al., 2013), much lower than the recommended level of 80%. Many factors including the number of trials, the number of subjects, and effect size

influence statistical power and reliability in ERP studies (Boudewyn et al., 2018; Clayson et al., 2019; Thigpen et al., 2017) and analysis of ERP datasets involves many researcher degrees of freedom and can lead to bogus significant ERP effects (Luck & Gaspelin, 2017). Thus, to conduct robust and replicable research using the CDA, it is critical to explicitly consider the number of subjects and trials that are needed to observe the effect of interest. This is the only way for CDA studies to provide a robust platform for theory-building in cognitive neuroscience.

To this end, we present estimates of statistical power at various numbers of trials and sample sizes for the presence of the CDA itself, as well as the presence of a set-size difference in the CDA. To achieve this, we conducted a subsampling analysis of two very large datasets that contain measures of CDA amplitudes at various set sizes; Hakim et al. (2019) and Unsworth et al. (2015). This approach enables the visualization of power contours that show the relationship between the number of subjects and trials and expected statistical power (Baker et al., 2020) and has also been recently applied by others (Adam et al., 2020; Von Gunten & Bartholow, 2019; Westfall et al., 2014; Xu et al., 2018). Further, we use these datasets to build predictive models to extrapolate the power beyond the limits of the aforementioned datasets. We supplement these results with an interactive app (https://williamngiam.shinyapps.io/CDAPower) that allows users to estimate their power to detect set-size differences in the CDA. This will allow researchers to optimize their parameters for their studies and justify their sample sizes.

2 | METHOD

2.1 | Datasets

The following datasets were selected as they contain EEG data on a lateralized change-detection task at multiple set sizes collected from a substantial number of subjects and trials —necessary to measure the CDA and reliably estimate statistical power.

2.1.1 | Unsworth et al. (2015)—The Unsworth et al. (2015) dataset contained 183 participants recruited from the University of Oregon community. The participants completed a battery of working memory and fluid intelligence tasks. Here, we analyzed the EEG data recorded during the lateralized color change-detection task. On each trial of the lateralized color change-detection task, participants were cued to one side of the screen. Two or six colored squares were then presented on both sides of the screen for 500 ms before a 900 ms retention interval. Each participant completed a total of 400 trials (200 at both set size 2 and set size 6). EEG data were recorded from 22 tin electrodes (Electro-Cap International, Eaton, OH) with impedance values kept below 10kΩ. The CDA was recorded at the International 10/20 sites PO3, PO4, O1, O2, OL, OR, T5 and T6, and was calculated by averaging the difference in amplitude between the contralateral and ipsilateral sides of the to-be-remembered hemifield of each electrode pair from 400 to 1,000 ms after the offset of the memory array. Trials were inspected and rejected if they contained eye movements, blinks, muscle noise, or blocking. We excluded subjects with fewer than 170 remaining trials per set-size condition after artifact rejection, leaving 135 participants for our analysis.

2.1.2 | Hakim et al. (2019)—We collapsed the data collected from 97 participants (73 unique) of the University of Chicago community across both experiments in the Hakim et al. (2019) dataset. Participants performed either a color change-detection task (in 2 sub-experiments; Experiment 1) or location change-detection task (in two sub-experiments; Experiment 2). On each trial, participants were cued to one side of the screen. Two or four stimuli—colored squares in Experiment 1 or circles in Experiment 2—were presented in both hemifields for 150 ms before a 1,300 ms retention period. As Hakim et al. (2019) found no effect of sub-experiment across the study, we collapsed the data across sub-experiments in these analyses as they did. EEG data were recorded from 30 active silver electrodes (actiCHamp, Brain Products, Munich, Germany) with impedance values kept below 10kΩ. The CDA was measured at the International 10/20 sites O1, O2, PO3, PO4, PO7, PO8, and P7, P8 for the color change-detection trials (Experiment 1) and the location change-detection trials (Experiment 2); 400 trials at set sizes 2 and 4 in both experiments. The CDA was calculated by taking the difference between the contralateral and ipsilateral electrodes relative to the to-be-remembered hemifield, then averaging across all trials and electrode pairs during the time window from 400 to 1,000 ms after the offset of the memory array. Trials were inspected and rejected if they contained eye movements, blinks, muscle noise, or blocking. We excluded subjects with fewer than 220 trials per set-size condition after artifact rejection, leaving 64 participants.

2.2 | Analysis

With each dataset, we conducted a subsampling analysis to estimate the statistical power at various combinations of trials and subjects to detect two effects—the presence of the CDA at the tested set sizes, and the difference in CDA amplitude between the tested set sizes. This entailed randomly sampling a given number of trials from a given number of subjects from the dataset and conducting a paired-samples *t* test on that subsampled data for the effects. At each combination of the number of trials and subjects, the proportion of statistically significant tests from 10,000 iterations was used as our estimate of statistical power.

We fit Bayesian multi-level models to each of the datasets to simulate data and estimate statistical power beyond the subject and trial bounds of the datasets. It should be noted that single-trial data distributions in both datasets were non-normal (there was statistically significant kurtosis in the distributions) but Gaussian models still provided overall accurate fits (see Results). All analysis code has been uploaded to Open Science Framework. Additionally, our results have been summarized using a Shiny app that can be used to estimate statistical power to detect the set-size effect.

3 | RESULTS

Estimated statistical power for each of the effects—the presence of the CDA at each set size, and the set-size difference in the CDA—is presented below as heatmaps. To anticipate our results, the presence of the CDA does not require very many clean trials to detect at standard sample sizes, but to detect CDA differences between set sizes requires a much larger number of trials and subjects.

3.1 | Unsworth et al. (2015)

Figure 1 shows the estimated statistical power to detect a significant overall CDA amplitude (i.e., significant differences in contralateral vs. ipsilateral activity) in the set size 2 (Mean = −0.42μV, *SD* = 0.44) and set size 6 (Mean = −0.56μV, *SD* = 0.45) conditions. Thus, to detect the *presence* of the CDA, 25 subjects and approximately 30 clean trials were needed per condition to achieve 80% power, a typical standard for a well-powered study (Cohen, 1988). However, it required substantially more subjects and trials to achieve acceptable statistical power to detect differences in CDA amplitude across set sizes (Mean Difference = −0.15μV, *SD* = 0.44) in CDA amplitude (i.e., set size 2 vs. set size 6). With the maximum number of clean trials per condition in our dataset (170 trials), at least 55 subjects were required to detect the set-size difference between 2 and 6 in the CDA (Figure 2).

Next, we generated a simulated dataset using Bayesian mixed modeling of the Unsworth et al. (2015) dataset to estimate power beyond the ranges of the dataset. Our simulated subsampling analysis did not meaningfully deviate from the subsampling of the actual dataset, RMSE = 0.0279 (Figure 2c), verifying the validity of the simulation. It should be noted that the simulated dataset slightly overestimates the number of subjects and trials required to achieve 80% power. As the single-trial distributions are non-normal (kurtosis > 3), this error in estimated statistical power may be improved by accounting for kurtosis. Going beyond the bounds of the Unsworth et al. (2015) dataset, for a typical set size of 25 subjects, 360 clean trials per condition would be needed to achieve an acceptable statistical power of 80% to detect the difference between set size 2 and set size 6 (Figure 3).

3.2 | Hakim et al. (2019) dataset

Figure 4 shows the estimated statistical power to detect a significant CDA amplitude in the set size 2 (Mean = −0.39μV *SD* = 0.44) and set size 4 (Mean = −0.56μV, *SD* = 0.56) conditions. With 25 subjects, a typical sample size for ERP experiments, approximately 40 to 50 clean trials are needed per condition for each condition to achieve 80% power for detecting the presence of the CDA component. Again, substantially more subjects and trials are needed to detect the set size difference in CDA, because this is a much smaller effect (Mean = −0.17μV, *SD* = 0.50). With the maximum number of clean trials per condition in the dataset (210 trials), at least 45 subjects were required to detect the set-size difference between 2 and 4 in the CDA (Figure 5a) with 80% power.

Like with the Unsworth et al. (2015) dataset, our simulated subsampling analysis did not meaningfully deviate from the subsampling of the actual dataset, RMSE = 0.0395, but we note that the simulated dataset slightly overestimates the number of subjects and trials required to achieve 80% power. As the single-trial distributions are non-normal (kurtosis > 3), this error in estimated statistical power may be improved by accounting for kurtosis. Going beyond the bounds of the Hakim et al. (2019) dataset, for a typical set size of 25 subjects, approximately 390 clean trials per condition would be needed to achieve an acceptable statistical power of 80% (Figure 6).

3.3 | Summary

With these analyses, we have estimated statistical power at various numbers of trials and subjects for three effects: the presence of the CDA, a significant difference in CDA between set sizes 2 and 6, and a significant difference in CDA between set sizes 2 and 4. The number of trials and subjects is clearly dependent on the effect size—to achieve enough power to robustly detect a larger effect like the presence of the CDA requires a smaller number of clean trials and subjects whereas with smaller effects like the set-size differences in the CDA, a significantly larger number of clean trials and subjects is required.

4 | DISCUSSION

As the CDA is a popular tool for studying VWM, it is important to design studies that achieve robust levels of statistical power for detecting the effects of interest. Underpowered studies contribute to false positives in the scientific literature (Maxwell, 2004). Reliance on past studies alone to guide sample sizes may have the undesired effect of perpetuating underpowered approaches. Thus, our goal was to provide a rigorous estimate of the number of subjects and trials required to achieve robust power by conducting subsampling analyses of two large EEG datasets collected during a change-detection task: Unsworth et al. (2015) and Hakim et al. (2019). These analyses show that to observe a significant CDA in any set-size condition (set sizes 2, 4, and 6 in our datasets here), approximately 50 clean trials for a typical sample size of 25 subjects is sufficient. However, to achieve the same statistical power of 80% for detecting set-size effects in the CDA requires larger sample sizes and number of trials than is typically seen in the literature. Our simulations suggest that closer to 400 clean trials with a sample size of 25 would be required to detect set-size differences in the CDA (at least between set size 2 and 4, and between set size 2 and 6). Our figures and interactive power calculator (https://williamngiam.shinyapps.io/CDAPower) provide researchers a principled method of selecting numbers of subjects and trials to include in their CDA experimental design, or to estimate the statistical power of their existing designs.

Our simulations demonstrate that the number of subjects and trials required to achieve robust power depends critically on the magnitude of the expected effect to be observed with the CDA. Here we quantify that for two classes of effects with different magnitudes—the presence of the CDA itself and set-size differences in CDA amplitude. On one end, the effect size for the CDA itself is large enough to make it fairly easy to reliably detect it, whereas the effect size for the set-size difference in the CDA (at least set size 2 vs. set size 6 and set size 2 vs. set size 4 comparisons) is much smaller and requires substantially larger sample sizes and clean trials collected. This may motivate changes in experimental designs such as including a set size 1 condition instead (if the goal of the study allows) as the magnitude of the CDA difference will be much larger when comparing to a set size 1 condition and thus will require fewer trials or subjects to achieve the same statistical power for detecting a set-size difference (see Fukuda et al., 2015 for a set-size function of CDA amplitudes). However, when the inclusion of a set size 1 condition may not be possible or is not informative to the research aims, it may be most expedient to boost power by increasing the number of trials collected rather than the number of subjects because collecting EEG data take a considerable investment of resources and time per subject from preparing the

electrode cap and organizing the subject for the experiment. Note that this recommendation is not applicable when the goal is to document individual differences in CDA activity, where a higher number of subjects will also be essential once an adequate number of trials has been collected.

Our power calculator (https://williamngiam.shinyapps.io/CDAPower) will help users estimate a sufficient number of trials and subjects required to detect various effect sizes with the CDA. This interactive power calculator (https://shiny.york.ac.uk/powercontours/) is also available for researchers to optimize the number of trials or subjects for their design given an expected effect size outside of the ones we have estimated here (Baker et al., 2020). Our estimations of statistical power serve as a guideline and may not generalize precisely to other CDA experiments that examine effects beyond set size and vary in significant ways to the datasets we have reanalyzed here. That is, researchers examining condition effects other than set size should not automatically carryover the estimates for set-size differences that we have presented here for their own experiments. For example, both datasets we subsampled from used simple color stimuli in their change-detection tasks. The magnitude of any set-size effect in the CDA and the trial-to-trial noise may vary with the use of different stimuli, differences in task demands, and differences in data quality (e.g., electrode impedance, Kappenman & Luck, 2010; electrode type, Laszlo et al., 2014; Mathewson et al., 2017; overall data quality, Luck et al., 2019, etc.). Researchers should be mindful of their expected effect sizes and subsequent statistical power of their designs, perhaps collecting substantially more trials per subject or more subjects to ensure the robustness of their observed effects.

Funding information

National Institute of Mental Health, Grant/Award Number: R01MH087214

References

Adam KCS, Doss MK, Pabon E, Vogel EK, & de Wit H (2020). Δ9-Tetrahydrocannabinol (THC) impairs visual working memory performance: A randomized crossover trial. Neuropsychopharmacology, 45, 1807–1816. 10.1038/s41386-020-0690-3 [PubMed: 32386395]

Adam KCS, Vogel EK, & Awh E (2017). Clear evidence for item limits in visual working memory. Cognitive Psychology, 97, 79–97. 10.1016/j.cogpsych.2017.07.001 [PubMed: 28734172]

Asp I, Störmer VS, & Brady TF (2019). *Greater visual working memory capacity for visually-matched stimuli when they are recognized as meaningful* [Preprint]. PsyArXiv. 10.31234/osf.io/r6njf

Astle DE, Harvey H, Stokes M, Mohseni H, Nobre AC, & Scerif G (2014). Distinct neural mechanisms of individual and developmental differences in VSTM capacity: Mechanisms of *VSTM* capacity. Developmental Psychobiology, 56(4), 601–610. 10.1002/dev.21126 [PubMed: 23775219]

Baker DH, Vilidaite G, Lygo FA, Smith AK, Flack TR, Gouws AD, & Andrews TJ (2020). Power contours: Optimising sample size and precision in experimental psychology and human neuroscience. Psychological Methods. 10.1037/met0000337

Balaban H, & Luria R (2015). The number of objects determines visual working memory capacity allocation for complex items. NeuroImage, 119, 54–62. 10.1016/j.neuroimage.2015.06.051 [PubMed: 26119024]

Boudewyn MA, Luck SJ, Farrens JL, & Kappenman ES (2018). How many trials does it take to get a significant ERP effect? It depends. Psychophysiology, 55(6), e13049. 10.1111/psyp.13049 [PubMed: 29266241]

Brady TF, Störmer VS, & Alvarez GA (2016). Working memory is not fixed-capacity: More active storage capacity for real-world objects than for simple stimuli. Proceedings of the National Academy of Sciences, 113(27), 7459–7464. 10.1073/pnas.1520027113

Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, & Munafò MR (2013). Power failure: Why small sample size undermines the reliability of neuroscience. Nature Reviews Neuroscience, 14(5), 365–376. 10.1038/nrn3475 [PubMed: 23571845]

Carlisle NB, Arita JT, Pardo D, & Woodman GF (2011). Attentional templates in visual working memory. Journal of Neuroscience, 31(25), 9315–9322. 10.1523/JNEUROSCI.1097-11.2011 [PubMed: 21697381]

Clayson PE, Carbine KA, Baldwin SA, & Larson MJ (2019). Methodological reporting behavior, sample sizes, and statistical power in studies of event-related potentials: Barriers to reproducibility and replicability. Psychophysiology, 56(11), e13437. 10.1111/psyp.13437 [PubMed: 31322285]

Cohen J (1988). Statistical power analysis for the behavioral sciences (pp. 18–74). : Lawrence Erlbaum Associates. 10.4324/9780203771587

Cowan N (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. The Behavioral and Brain Sciences, 24(1), 87–114; discussion 114-185. 10.1017/S0140525X01003922 [PubMed: 11515286]

Drew T, Horowitz TS, Wolfe JM, & Vogel EK (2011). Delineating the neural signatures of tracking spatial position and working memory during attentive tracking. Journal of Neuroscience, 31(2), 659–668. 10.1523/JNEUROSCI.1339-10.2011 [PubMed: 21228175]

Duarte A, Hearons P, Jiang Y, Delvin MC, Newsome RN, & Verhaeghen P (2013). Retrospective attention enhances visual working memory in the young but not the old: An ERP study: Retrospective attention and aging. Psychophysiology, 50(5), 465–476. 10.1111/psyp.12034 [PubMed: 23445536]

Emrich SM, Al-Aidroos N, Pratt J, & Ferber S (2009). Visual search elicits the electrophysiological marker of visual working memory. PLoS One, 4(11), e8042. 10.1371/journal.pone.0008042 [PubMed: 19956663]

Emrich SM, Al-Aidroos N, Pratt J, & Ferber S (2010). Finding memory in search: The effect of visual working memory load on visual search. The Quarterly Journal of Experimental Psychology, 63(8), 1457–1466. 10.1080/17470218.2010.483768 [PubMed: 20603775]

Feldmann-Wüstefeld T, Vogel EK, & Awh E (2018). Contralateral delay activity indexes working memory storage, not the current focus of spatial attention. Journal of Cognitive Neuroscience, 30(8), 1–11. 10.1162/jocn_a_01271 [PubMed: 28891784]

Fukuda K, Vogel E, Mayr U, & Awh E (2010). Quantity, not quality: The relationship between fluid intelligence and working memory capacity. Psychonomic Bulletin & Review, 17(5), 673–679. 10.3758/17.5.673 [PubMed: 21037165]

Fukuda K, Woodman GF, & Vogel EK (2015). Individual differences in visual working memory capacity: Contributions of attentional control to storage. In Mechanisms of sensory working memory: Attention and performance (Vol. XXV, pp. 105–119). Elsevier. 10.1016/B978-0-12-801371-7.00009-0

Göddertz A, Klatt L-I, Mertes C, & Schneider D (2018). Retroactive attentional shifts predict performance in a working memory task: Evidence by lateralized EEG patterns. Frontiers in Human Neuroscience, 12, 428. 10.3389/fnhum.2018.00428 [PubMed: 30405380]

Gunseli E, Meeter M, & Olivers CNL (2014). Is a search template an ordinary working memory? Comparing electrophysiological markers of working memory maintenance for visual search and recognition. Neuropsychologia, 60, 29–38. 10.1016/j.neuropsychologia.2014.05.012 [PubMed: 24878275]

Hakim N, Adam KCS, Gunseli E, Awh E, & Vogel EK (2019). Dissecting the neural focus of attention reveals distinct processes for spatial attention and object-based storage in visual working memory. Psychological Science, 30(4), 526–540. 10.1177/0956797619830384 [PubMed: 30817220]

Heuer A, & Schubö A (2016). The focus of attention in visual working memory: Protection of focused representations and its individual variation. PLoS One, 11(4), e0154228. 10.1371/journal.pone.0154228 [PubMed: 27099938]

Ikkai A, McCollough AW, & Vogel EK (2010). Contralateral delay activity provides a neural measure of the number of representations in visual working memory. Journal of Neurophysiology, 103(4), 1963–1968. 10.1152/jn.00978.2009 [PubMed: 20147415]

Ioannidis JPA (2005). Why most published research findings are false. PLoS Medicine, 2(8), e124. 10.1371/journal.pmed.0020124 [PubMed: 16060722]

Jost K, Bryck RL, Vogel EK, & Mayr U (2011). Are old adults just like low working memory young adults? Filtering efficiency and age differences in visual working memory. Cerebral Cortex, 21(5), 1147–1154. 10.1093/cercor/bhq185 [PubMed: 20884722]

Kappenman ES, & Luck SJ (2010). The effects of electrode impedance on data quality and statistical significance in ERP recordings. Psychophysiology, 47(5), 888–904. 10.1111/j.1469-8986.2010.01009.x [PubMed: 20374541]

Kuo B-C, Stokes MG, & Nobre AC (2012). Attention modulates maintenance of representations in visual short-term memory. Journal of Cognitive Neuroscience, 24(1), 51–60. 10.1162/jocn_a_00087 [PubMed: 21736457]

Laszlo S, Ruiz-Blondet M, Khalifian N, Chu F, & Jin Z (2014). A direct comparison of active and passive amplification electrodes in the same amplifier system. Journal of Neuroscience Methods, 235, 298–307. 10.1016/j.jneumeth.2014.05.012 [PubMed: 25075801]

Lee E-Y, Cowan N, Vogel EK, Rolan T, Valle-Inclan F, & Hackley SA (2010). Visual working memory deficits in patients with Parkinson's disease are due to both reduced storage capacity and impaired ability to filter out irrelevant information. Brain, 133(9), 2677–2689. 10.1093/brain/awq197 [PubMed: 20688815]

Leonard CJ, Kaiser ST, Robinson BM, Kappenman ES, Hahn B, Gold JM, & Luck SJ (2013). Toward the neural mechanisms of reduced working memory capacity in schizophrenia. Cerebral Cortex, 23(7), 1582–1592. 10.1093/cercor/bhs148 [PubMed: 22661407]

Luck SJ, & Gaspelin N (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). Psychophysiology, 54(1), 146–157. 10.1111/psyp.12639 [PubMed: 28000253]

Luck SJ, Stewart AX, Simmons AM, & Rhemtulla M (2019). *Standardized measurement error as a universal measure of data quality for event-related potentials: An overview* [Preprint]. PsyArXiv. 10.31234/osf.io/jc3sd

Luck SJ, & Vogel EK (1997). The capacity of visual working memory for features and conjunctions. Nature, 390(6657), 279–281. 10.1038/36846 [PubMed: 9384378]

Luck SJ, & Vogel EK (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. Trends in Cognitive Sciences, 17(8), 391–400. 10.1016/j.tics.2013.06.006 [PubMed: 23850263]

Luria R, Balaban H, Awh E, & Vogel EK (2016). The contralateral delay activity as a neural measure of visual working memory. Neuroscience & Biobehavioral Reviews, 62, 100–108. 10.1016/j.neubiorev.2016.01.003 [PubMed: 26802451]

Mathewson KE, Harrison TJL, & Kizuk SAD (2017). High and dry? Comparing active dry EEG electrodes to active and passive wet electrodes: Active dry vs. active & passive wet EEG electrodes. Psychophysiology, 54(1), 74–82. 10.1111/psyp.12536 [PubMed: 28000254]

Maxwell SE (2004). The persistence of underpowered studies in psychological research: Causes, consequences, and remedies. Psychological Methods, 9(2), 147–163. 10.1037/1082-989X.9.2.147 [PubMed: 15137886]

Meconi F, Luria R, & Sessa P (2014). Individual differences in anxiety predict neural measures of visual working memory for untrustworthy faces. Social Cognitive and Affective Neuroscience, 9(12), 1872–1879. 10.1093/scan/nst189 [PubMed: 24493843]

Pagano S, Fait E, Monti A, Brignani D, & Mazza V (2015). Electrophysiological correlates of subitizing in healthy aging. PLoS One, 10(6), e0131063. 10.1371/journal.pone.0131063 [PubMed: 26098959]

Prime DJ, & Jolicoeur P (2009). Mental rotation requires visual short-term memory: Evidence from human electric cortical activity. Journal of Cognitive Neuroscience, 22(11), 2437–2446. 10.1162/jocn.2009.21337

Qi S, Ding C, & Li H (2014). Neural correlates of inefficient filtering of emotionally neutral distractors from working memory in trait anxiety. Cognitive, Affective, & Behavioral Neuroscience, 14(1), 253–265. 10.3758/s13415-013-0203-5

Quirk C, Adam KCS, & Vogel EK (2020). No evidence for an object working memory capacity benefit with extended viewing time. eNeuro, 7(5), ENEUR0.0150-20.2020. 10.1523/ENEURO.0150-20.2020

Rabbitt LR, Roberts DM, McDonald CG, & Peterson MS (2017). Neural activity reveals perceptual grouping in working memory. International Journal of Psychophysiology, 113, 40–45. 10.1016/j.ijpsycho.2017.01.005 [PubMed: 28088351]

Salahub C, & Emrich S (2019). *Fear not! Anxiety biases attentional enhancement of threat without impairing working memory filtering* [Preprint]. PsyArXiv. 10.31234/osf.io/rn4fa

Salahub C, Lockhart HA, Dube B, Al-Aidroos N, & Emrich SM (2019). Electrophysiological correlates of the flexible allocation of visual working memory resources. Scientific Reports, 9(1), 19428. 10.1038/s41598-019-55948-4 [PubMed: 31857657]

Sander MC, Werkle-Bergner M, & Lindenberger U (2011). Contralateral delay activity reveals life-span age differences in top-down modulation of working memory contents. Cerebral Cortex, 21(12), 2809–2819. 10.1093/cercor/bhr076 [PubMed: 21527784]

Smaldino PE, & McElreath R (2016). The natural selection of bad science. Royal Society Open Science, 3(9), 160384. 10.1098/rsos.160384 [PubMed: 27703703]

Spronk M, Vogel EK, & Jonkman LM (2012). Electrophysiological evidence for immature processing capacity and filtering in visuospatial working memory in adolescents. PLoS One, 7(8), e42262. 10.1371/journal.pone.0042262 [PubMed: 22927923]

Spronk M, Vogel EK, & Jonkman LM (2013). No behavioral or ERP evidence for a developmental lag in visual working memory capacity or filtering in adolescents and adults with ADHD. PLoS One, 8(5), e62673. 10.1371/journal.pone.0062673 [PubMed: 23650525]

Stout DM, Shackman AJ, & Larson CL (2013). Failure to filter: Anxious individuals show inefficient gating of threat from working memory. Frontiers in Human Neuroscience, 7, 58. 10.3389/fnhum.2013.00058 [PubMed: 23459454]

Thigpen NN, Kappenman ES, & Keil A (2017). Assessing the internal consistency of the event-related potential: An example analysis: Assessing internal consistency of the ERP. Psychophysiology, 54(1), 123–138. 10.1111/psyp.12629 [PubMed: 28000264]

Unsworth N, Fukuda K, Awh E, & Vogel EK (2015). Working memory delay activity predicts individual differences in cognitive abilities. Journal of Cognitive Neuroscience, 27(5), 853–865. 10.1162/jocn_a_00765 [PubMed: 25436671]

Vogel EK, & Machizawa MG (2004). Neural activity predicts individual differences in visual working memory capacity. Nature, 428(6984), 748–751. 10.1038/nature02447 [PubMed: 15085132]

Vogel EK, McCollough AW, & Machizawa MG (2005). Neural measures reveal individual differences in controlling access to working memory. Nature, 438(7067), 500–503. 10.1038/nature04171 [PubMed: 16306992]

Von Gunten CD, & Bartholow BD (2019). Exploring power in response inhibition tasks using the bootstrap: The impact of number of participants, number of trials, effect magnitude, and study design. International Journal of Psychophysiology, S0167876019304726. 10.1016/j.ijpsycho.2019.08.008

Westfall J, Kenny DA, & Judd CM (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. Journal of Experimental Psychology: General, 143(5), 2020–2045. 10.1037/xge0000014 [PubMed: 25111580]

Wiegand I, Hennig-Fast K, Kilian B, Müller HJ, Töllner T, Möller H-J, Engel RR, & Finke K (2016). EEG correlates of visual short-term memory as neuro-cognitive endophenotypes of ADHD. Neuropsychologia, 85, 91–99. 10.1016/j.neuropsychologia.2016.03.011 [PubMed: 26972967]

Williams LH, & Drew T (2020). Maintaining rejected distractors in working memory during visual search depends on search stimuli: Evidence from contralateral delay activity. Attention, Perception, & Psychophysics. 10.3758/s13414-020-02127-7

Williams M, & Woodman GF (2012). Directed forgetting and directed remembering in visual working memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 38(5), 1206–1220. 10.1037/a0027389

Woodman GF, & Arita JT (2011). Direct electrophysiological measurement of attentional templates in visual working memory. Psychological Science, 22(2), 212–215. 10.1177/0956797610395395 [PubMed: 21193780]

Xie W, & Zhang W (2017). Familiarity speeds up visual short-term memory consolidation. Journal of Experimental Psychology: Human Perception and Performance, 43(6), 1207–1221. 10.1037/xhp0000355 [PubMed: 28287761]

Xu Z, Adam KCS, Fang X, & Vogel EK (2018). The reliability and stability of visual working memory capacity. Behavior Research Methods, 50(2), 576–588. 10.3758/s13428-017-0886-6 [PubMed: 28389852]

Zaehle T, Becke A, Naue N, Machts J, Abdulla S, Petri S, Kollewe K, Dengler R, Heinze H-J, Vielhaber S, & Müller NG (2013). Working memory in ALS patients: Preserved performance but marked changes in underlying neuronal networks. PLoS One, 8(8), e71973. 10.1371/journal.pone.0071973 [PubMed: 23951274]

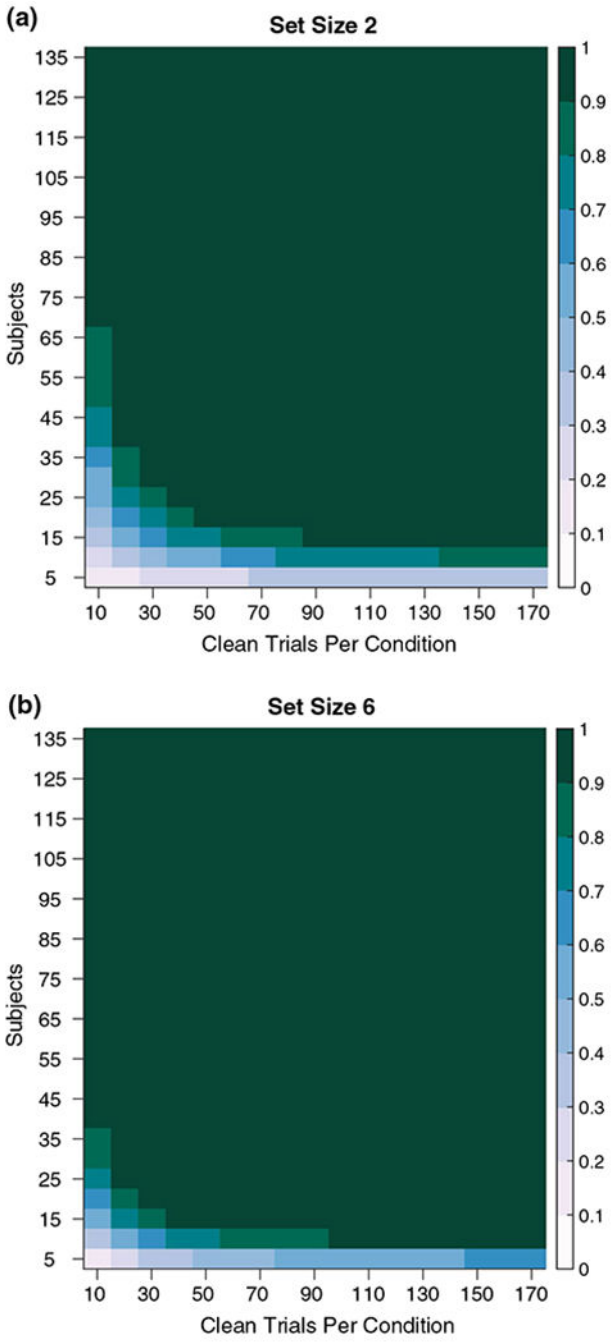

FIGURE 1.

(a) Estimated statistical power for observing a significant CDA at set size 2. Number of subjects refers to the subsampled sample size and the number of clean trials per condition refers to the number of trials following artifact rejection that were sampled. Each combination of the number of subjects and clean trials per condition is plotted as a cell in the heatmap with its estimated statistical power portrayed by color on the accompanying color scale. (b) Estimated statistical power for observing a significant CDA at set size 6

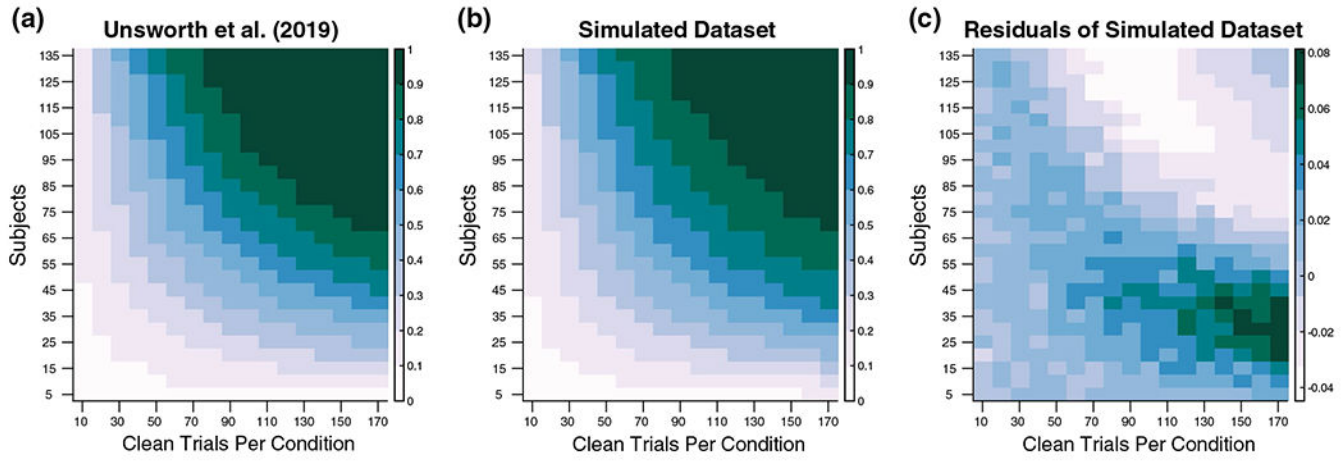

FIGURE 2.

(a) Estimated statistical power for observing a significant difference in CDA amplitude between set sizes 2 and 6. (b) Simulated statistical power for observing a significant difference in CDA between set sizes 2 and 6 (c) The residual of statistical power estimated from the actual dataset and estimated from the simulated dataset (actual – simulated)

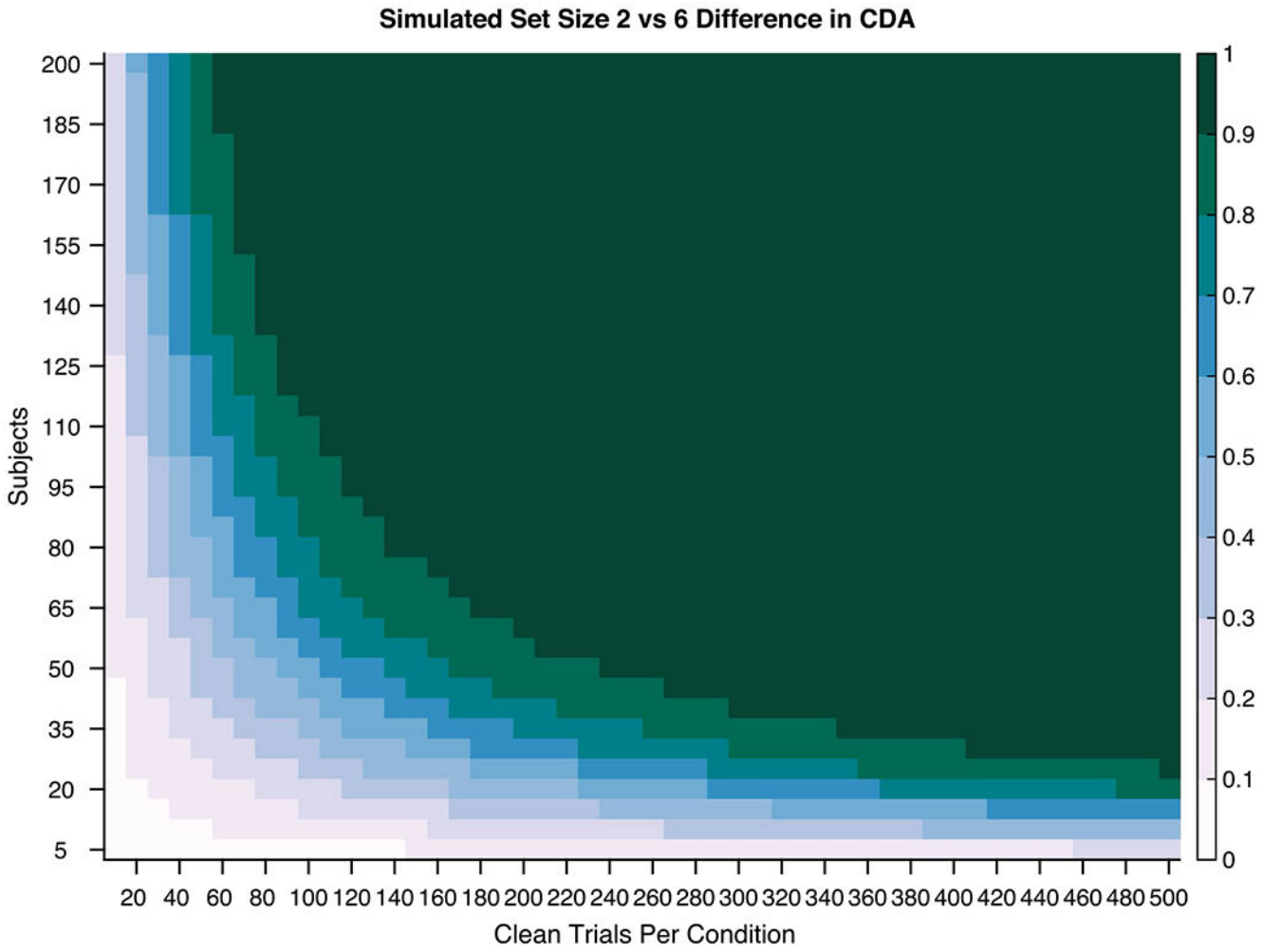

FIGURE 3.

Estimated statistical power for observing a significant difference in CDA amplitude between set sizes 2 and 6 from our simulated models

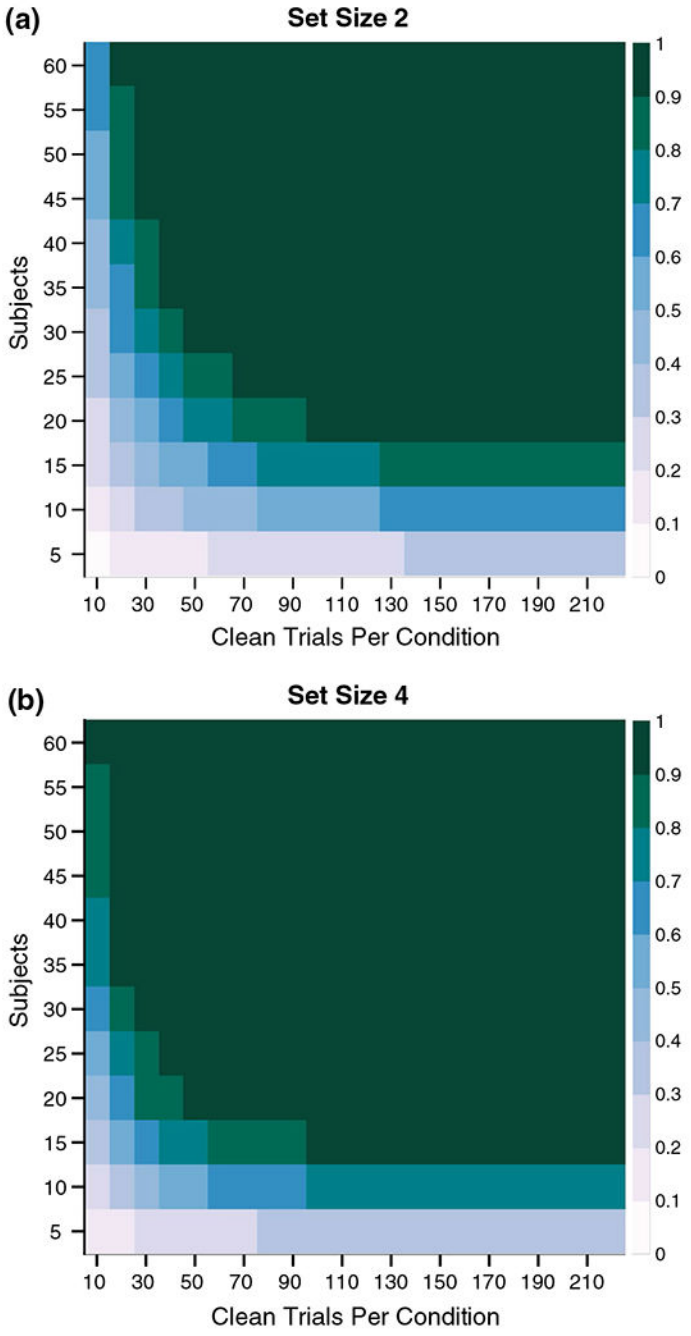

FIGURE 4.

(a) Estimated statistical power for observing a significant CDA at set size 2. (b) Estimated statistical power for observing a significant CDA at set size 4

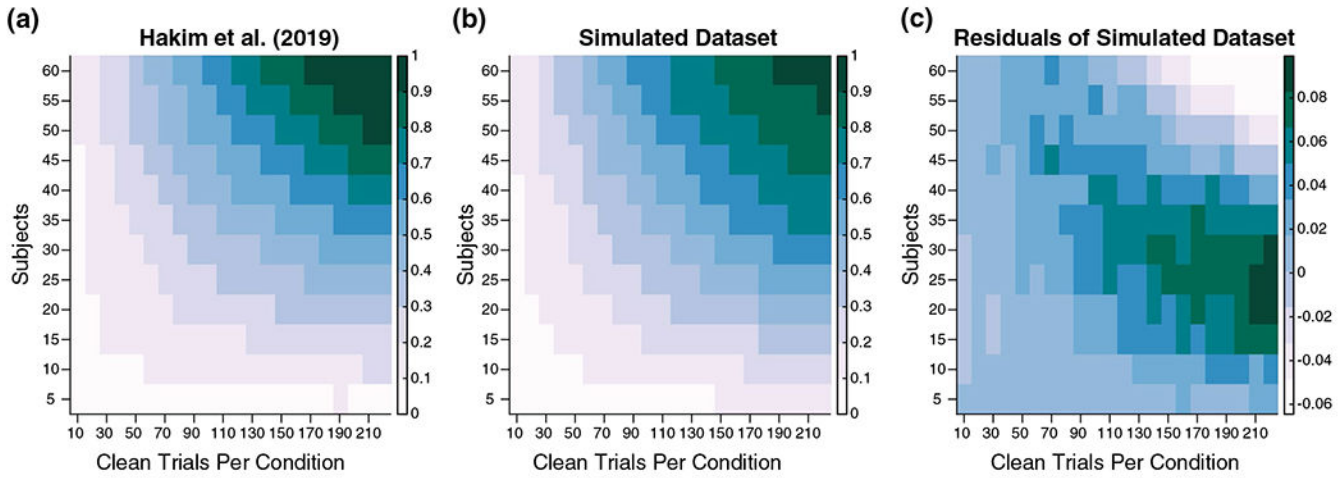

FIGURE 5.

(a) Estimated statistical power for observing a significant difference in CDA amplitude between set sizes 2 and 4. (b) Simulated statistical power for observing the same effect. (c) The residual of estimated statistical power from the simulated dataset compared to the Hakim et al. (2019) dataset (actual – simulated)

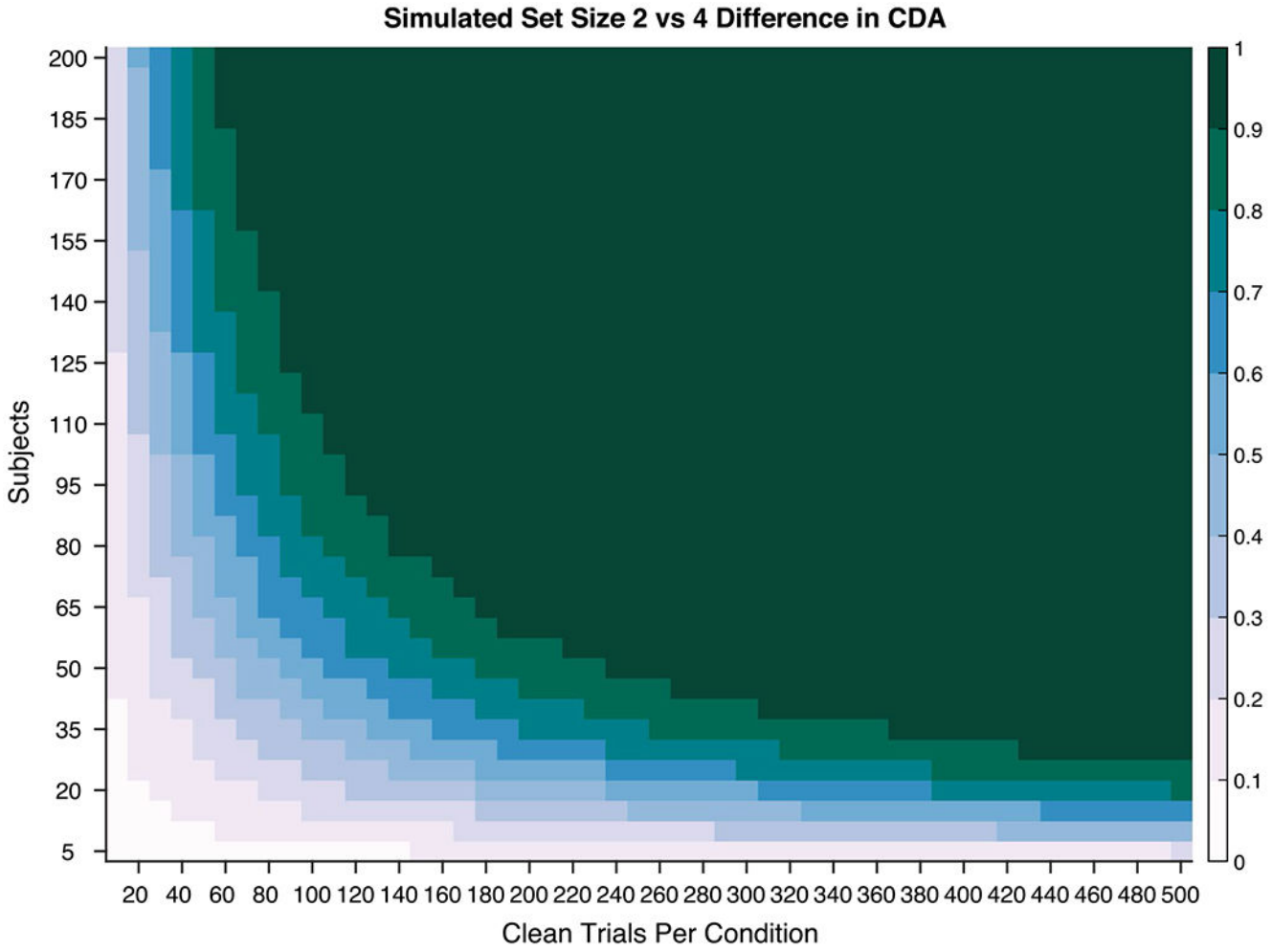

FIGURE 6.

Simulated statistical power for observing a significant difference in CDA amplitude between set sizes 2 and 4 beyond the bounds of the Hakim et al. (2019) dataset