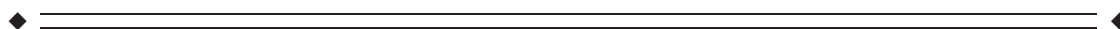# Meta-Analysis of Psychophysiological Interactions: Revisiting Cluster-Level Thresholding and Sample Sizes

**David V. Smith[1]* and Mauricio R. Delgado[2]***

[1]*Department of Psychology, Temple University, Philadelphia, Pennsylvania, 19122*
[2]*Department of Psychology, Rutgers University, Newark, New Jersey, 17102*

Abstract: Within the neuroimaging community, coordinate based meta-analyses (CBMAs) are essential for aggregating findings across studies and testing whether those studies report similar anatomical locations. This approach has been predominantly applied to studies that focus on whether activation of a brain region is associated with a given psychological process. In a recent paper, we used CBMA to examine a distinct set of studies—that is, those focusing on whether connectivity between brain regions is modulated by a given psychological process (Smith et al. [2016]: Hum Brain Mapp 37:2904–2917). Specifically, we reviewed 284 studies examining brain connectivity with psychophysiological interactions (PPI). Our meta-analytic results indicated that PPI yields connectivity patterns that are consistent across studies and that can be specific for a given psychological process and seed region. After publication of our findings, we learned that the analysis software we used to conduct our CBMAs (GingerALE v2.3.3) contained an implementation error that led to results that were more liberal than intended. Here, we comment on the impact of this implementation error on the results of our paper, new recommendations for sample sizes in CBMAs, and the importance of communication between software users and developers. We show that our key claims are supported in a reanalysis and that our results are robust to new guidelines on sample sizes. *Hum Brain Mapp 38:588–591, 2017.* © **2016 Wiley Periodicals, Inc.**

Key words: psychophysiological interaction; PPI; CBMA; meta-analysis; fMRI; open science
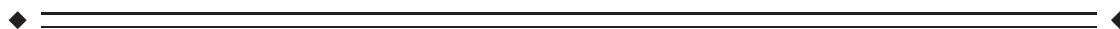
INTRODUCTION

Neuroimaging results are conventionally reported in terms of coordinates within a 3-dimensional stereotactic system. This reporting system motivated the development of coordinate-based meta-analyses (CBMAs), which allow for the quantification of consistency and specificity across studies [Yarkoni et al., 2010]. Although the CBMA approach has largely been applied to studies examining how activation of a region (a coordinate) is associated with a given psychological process, it has also been used to evaluate whether regions are consistently co-activated [Pauli et al., 2016; Robinson et al., 2012]. This approach—called meta-analytic connectivity modeling [Robinson et al., 2010]—is analogous to functional connectivity and therefore suffers from similar limitations [Gerstein and Perkel, 1969]. Specifically, changes in functional connectivity between two regions can arise due to changes in signal or noise in either region, or via a change in connectivity with another region [Friston, 2011].

In recent study, we addressed this problem by conducting CMBAs on studies examining psychophysiological interactions (PPI), a popular brain connectivity analysis approach that can be interpreted as a simple (linear) model of effective connectivity [Smith et al., 2016]. PPI evaluates whether an interaction between a seed region and psychological context is expressed in a target region [Friston et al., 1997]. For example, researchers who select an amygdala seed region and an emotional task may find a target region in medial prefrontal cortex (MPFC), indicating that, during emotional processing, the amygdala contributes to the response in the MPFC. We examined the reported target regions from 284 PPI studies using different seed regions and psychological contexts—a strategy that allowed us to quantify the consistency and specificity of PPI results. We conducted our analyses using Ginger-ALE (v2.3.3), a powerful and easy-to-use program for CBMAs. Our analyses supported two broad conclusions. First, PPI studies produce reliable results, consistently identifying similar targeting regions across studies. For instance, PPI studies using the amygdala as a seed region and emotion as the psychological context reliably reported target regions within the MPFC. Notably, this particular result has already been replicated and extended by an independent group [Di et al., 2016]. Second, PPI studies can produce results that are highly specific to a given seed region and psychological context. For example, PPI studies using the dorsolateral prefrontal cortex (DLPFC) as a seed region identified targets in the posterior cingulate cortex when the psychological context involved cognitive control and targets in the amygdala when the psychological context did not involve cognitive control. These results highlight both the consistency and specificity of PPI studies [Smith et al., 2016].

Shortly after our paper appeared online, we learned that the version of GingerALE we used contained an implementation error in the cluster-level correction code, which resulted in thresholds that did not control the family wise error (FWE) rate. This implementation error is detailed in a recent Technical Report authored by the developers of GingerALE [Eickhoff et al., 2016a]. The developers recommend that users repeat their analyses with the newest version of GingerALE (v2.3.6) and compare the results to those in their original reports. Depending on the impact of the implementation error, the developers also discuss different approaches for corrective communications. For example, minimal corrections may only require a comment on Pubmed confirming the original results while more substantive corrections may need to be outlined and discussed in a Comment-type article [Eickhoff et al., 2016a].

Based on the developers' recommendations, the purpose of this article is to discuss the impact of the GingerALE's implementation error on our results. Our reanalysis utilized a conventional cluster-level threshold of $P = 0.05$ with cluster-forming threshold of $P = 0.001$ [Eickhoff et al., 2016b]. Importantly, the reanalysis confirmed all of our

key claims. We note, however, that the reanalysis also indicated that three targets reported in our original results did not survive our specified FWE rate (see targets below). Ignoring these three targets in our paper does not alter any of our conclusions. Of these regions, only the inferior lateral occipital cortex (iLOC) target was mentioned in our Discussion section.

1. dorsolateral prefrontal cortex (DLPFC) target in Figure 3B (i.e., DLPFC seed with cognitive control studies);
2. inferior lateral occipital cortex (iLOC) target in Figure 3C (i.e., amygdala seed with emotion studies);
3. ventrolateral prefrontal cortex (VLPFC) target in Supporting Information Figure 2 (i.e., ventral striatum seed contrasted against other seeds).

In addition to these specific changes, we also note two additional features of our reanalysis. First, one of our larger clusters (superior temporal sulcus in Figure 3D) was only evident with a cluster-forming threshold of $P = 0.005$. Although higher cluster-forming threshold (e.g., $P = 0.001$) are important for parametric analyses using random field theory, recent large-scale simulations have suggested that nonparametric analyses (i.e., permutation-based testing) can maintain specified FWE rates in conventional fMRI analyses, even with relatively low cluster-forming thresholds (e.g., $P = 0.01$) [Eklund et al., 2016]. While GingerALE employs permutation-based testing in its cluster-extent thresholding, it remains unclear whether FWE rates become inflated at lower cluster-forming thresholds, so users are advised to use $P = 0.001$ [Eickhoff et al., 2016a]. Second, our reanalysis utilized a cluster-level threshold that was more liberal than our original paper ($P = 0.00625$). This original threshold was intended to add an extra layer of protection against multiple comparisons (i.e., four bidirectional contrasts); however, it is quite rare to see corrected thresholds that are more stringent than $P = 0.05$, likely because such thresholds elevate the risk of Type 2 errors [Lieberman and Cunningham, 2009]. Indeed, all of the changes that stem from our reanalysis reflect a tradeoff between Type 1 and Type 2 errors, and our goal here is to provide accurate information about the nature of that tradeoff within our study. Interested readers are welcome to review our new statistical maps on the NeuroVault repository: http://neurovault.org/collections/1406 [Gorgolewski et al., 2015].

We have also revisited another aspect of our paper because of a recent development in sample size recommendations for CBMAs. Sample sizes (i.e., number of experiments) are an important issue within CBMAs using activation likelihood estimation (ALE). Indeed, when an ALE score is based on a small number of experiments, there is considerable risk that the observed results can be driven by a single experiment—an effect that would obviously undermine the value of the meta-analysis. Our paper

relied on previous anecdotal recommendations for sample size, which stated that "at least 10–15 experiments" are needed to reduce the likelihood that meta-analytic results are driven by a single experiment [Eickhoff and Bzdok, 2013]. In a recent study, this issue was examined quantitatively using massive sets of simulations [Eickhoff et al., 2016b]. This study came to the conclusion that a sample size of at least 17–20 experiments is necessary to prevent (a) one experiment from accounting for more than 50% of the ALE score and (b) two experiments from accounting for more than 80% of the ALE score. Based on these recent results, we revisited sample size in two of our analyses that utilized fewer than 17 experiments. In Figure 3A, we reported two clusters that were based on 10 experiments. Within the fusiform face area (FFA) cluster, we found that four experiments contributed to the result, each accounting for 16.84%–33.15% of the ALE score (the top two experiments accounted for 63.71%). And within the primary somatosensory (S1) cluster, we found that five experiments contributed to the result, each accounting for 6.64%–37.15% of the ALE score (the top two experiments accounted for 67.18%). We repeated this analysis for Figure 3B, which reported a cluster in posterior cingulate cortex (PCC) using 15 experiments. We found that five experiments contributed to the result, each accounting for 0.27%–42.29% of the ALE score (the top two experiments accounted for 66.26%). Taken together, these observations suggest that our original findings were not biased by small sample sizes.

Analytical tools within the neuroimaging community and recommendations for best practices are continually being improved. How can these advances be communicated to users? Although the there is a wide range of platforms for scholarly communication, we highlight three strategies that may help with software-specific developments. First, in cases where an analysis program is affected by an implementation error, the developers can describe the error and discuss the need for reanalysis. This strategy is exemplified and discussed in the Technical Note from the GingerALE developers [Eickhoff et al., 2016a]. Second, it would be helpful for developers to announce software updates via email. For example, the FMRIB Software Library (FSL) analysis package for fMRI data encourages all users to provide an email address when downloading FSL [Jenkinson et al., 2012]. Whenever there is an update to the software, all users are notified immediately. Third, we encourage users to be actively engaged in support forums because of the tremendous benefits associated with open communication and dialogue between users and developers. We realize that some of these forums have high traffic (e.g., dozens of emails each day) and thus may be daunting to read carefully or regularly; however, they provide an excellent resource for learning about new techniques and solutions to common problems. In addition, active engagement with a support forum can sometimes provide an early warning signal that there may be a peculiar behavior in a program. Indeed, the possibility of a problem with the cluster-level thresholding option in GingerALE was first raised on the support forums. Had we been engaged with that support forum, we likely would have noticed the discussion and delayed the submission of our article accordingly. We hope these points illustrate the importance of communication between users and developers, while also highlighting the advantages of using software and tools that promote dialogue.

In summary, we have demonstrated that the implementation error in GingerALE did not affect the original conclusions of our PPI meta-analysis. We have also shown that our results are robust to revised recommendations for minimum sample sizes in CBMAs. Although CBMAs are currently the primary tool for assessing the consistency and specificity of neuroimaging results, image-based meta-analyses [Salimi-Khorshidi et al., 2009] may eventually rise in popularity as more authors share unthresholded statistical maps and adopt more open science practices [Gorgolewski and Poldrack, 2016; McKiernan et al., 2016]. Within the context of PPI studies, image-based meta-analyses would have the power to detect subthreshold connectivity patterns across studies, thus improving our understanding of how brain connectivity shapes behavior.

ACKNOWLEDGMENTS

REFERENCES

Di X, Huang J, Biswal BB (2016): Task modulated brain connectivity of the amygdala: A meta-analysis of psychophysiological interactions. Brain Struct Funct. http://www.ncbi.nlm.nih.gov/pubmed/27259584.

Eickhoff SB, Bzdok D (2013): Meta-analyses in basic and clinical neuroscience: State of the art and perspective. In: Ulmer S, Jansen O, editors. Fmri: Basics and Clinical Applications. Berlin, Heidelberg: Springer Berlin Heidelberg. pp. 77–87.

Eickhoff SB, Laird A, Fox PM, Lancaster JL, Fox PT (2016a): implementation errors in the gingerale software: Description and recommendations. Hum Brain Mapp. http://www.ncbi.nlm.nih.gov/pubmed/27511454.

Eickhoff SB, Nichols TE, Laird AR, Hoffstaedter F, Amunts K, Fox PT, Bzdok D, Eickhoff CR (2016b): Behavior, sensitivity, and power of activation likelihood estimation characterized by massive empirical simulation. Neuroimage 137:70–85.

Eklund A, Nichols TE, Knutsson H (2016): Cluster failure: Why fmri inferences for spatial extent have inflated false-positive rates. Proc Natl Acad Sci U S A 113:7900–7905.

Friston KJ (2011): Functional and effective connectivity: A review. Brain Connect 1:13–36.

Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ (1997): Psychophysiological and modulatory interactions in neuroimaging. Neuroimage 6:218–229.

Gerstein GL, Perkel DH (1969): Simultaneously recorded trains of action potentials: Analysis and functional interpretation. Science 164:828–830.

Gorgolewski KJ, Poldrack RA (2016): A practical guide for improving transparency and reproducibility in neuroimaging research. PLoS Biol 14:e1002506.

Gorgolewski KJ, Varoquaux G, Rivera G, Schwarz Y, Ghosh SS, Maumet C, Sochat VV, Nichols TE, Poldrack RA, Poline JB, Yarkoni T, Margulies DS (2015): Neurovault.org: A web-based repository for collecting and sharing unthresholded statistical maps of the human brain. Front Neuroinform 9:8.

Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM (2012): Fsl. Neuroimage 62:782–790.

Lieberman MD, Cunningham WA (2009): Type I and Type Ii error concerns in fmri research: Re-balancing the scale. Social Cogn Affect Neurosci 4:423–428.

McKiernan EC, Bourne PE, Brown CT, Buck S, Kenall A, Lin J, McDougall D, Nosek BA, Ram K, Soderberg CK, Spies JR, Thaney K, Updegrove A, Woo KH, Yarkoni T (2016): How open science helps researchers succeed. Elife 5:10.

Pauli WM, O'Reilly RC, Yarkoni T, Wager TD (2016): Regional specialization within the human striatum for diverse psychological functions. Proc Natl Acad Sci U S A, 113:1907–1912.

Robinson JL, Laird AR, Glahn DC, Lovallo WR, Fox PT (2010): Metaanalytic connectivity modeling: Delineating the functional connectivity of the human amygdala. Hum Brain Mapp 31: 173–184.

Robinson JL, Laird AR, Glahn DC, Blangero J, Sanghera MK, Pessoa L, Fox PM, Uecker A, Friehs G, Young KA, Griffin JL, Lovallo WR, Fox PT (2012): The functional connectivity of the human caudate: An application of meta-analytic connectivity modeling with behavioral filtering. Neuroimage 60:117–129.

Salimi-Khorshidi G, Smith SM, Keltner JR, Wager TD, Nichols TE (2009): Meta-analysis of neuroimaging data: A comparison of image-based and coordinate-based pooling of studies. Neuroimage 45:810–823.

Smith DV, Gseir M, Speer ME, Delgado MR (2016): Toward a cumulative science of functional integration: A meta-analysis of psychophysiological interactions. Hum Brain Mapp 37: 2904–2917.

Yarkoni T, Poldrack RA, Van Essen DC, Wager TD (2010): Cognitive neuroscience 2.0: building a cumulative science of human brain function. Trends Cogn Sci 14:489–496. 10.1016/j.tics.2010.08.004