SM3: Identifying the "main claim/finding" for each study

Peder M. Isager

9/17/2021

Some information that is relevant for assessing replication value is related to individual empirical findings within studies (e.g., the precision of the estimate for a particular effect size of interest). If we want to use such information to compare the replication value of two studies, we first need to decide which findings from each study to use for our comparison. For example, consider the implication of using the standard error of the mean to calculate RV_{Cn}. If we do not approximate the standard error via the total sample size of the study, standard error is related to a particular mean estimate within the study. Since a study may report many mean estimates, it may be related to any number of standard errors. Thus, it is no longer enough to decide which studies to include as replication candidates. We now also have to decide which specific findings from these studies to consider, because RV_{Cn} estimates depend on statistical information from these findings.

We conducted a study to try and identify the *main finding* of each study in our set of replication candidates. The main finding was defined as the reported finding which is centrally highlighted in either the abstract or conclusion section of the article in which the study is reported, and which seemed to be the focus point of the study design. For example, the finding that the fusiform face area is reliably and selectively activated by images of faces is the main finding used to support the more general claim that faces are processed in a specific spatial location within the human brain. The ultimate goal of this study, in conjunction with the pilot study reported in supplementary material SM2, was to identify indicators of statistical uncertainty for each main finding (such as standard error of the mean, and Bayesian posterior evidence) from which different estimators of replication value could be constructed, calculated, and compared with RV_{Cn}.

Main findings for each paper had to be coded manually. We developed a general coding procedure, instructing coders on where in the paper to look for mentions of the main finding, and what would indicate that something is a main finding. Three co-authors (PMI, AvtV, LG) then applied this procedure to a small set of studies within our candidate set to test the feasibility of the coding effort. All data and materials from this small coding effort is available on OSF (https://osf.io/953du/). Below follows a brief summary of our own conclusions.

Our pilot suggested that main findings from each study could indeed be identified. Identification was relatively time-intensive. A main finding could be identified within a few minutes on average, but overall time taken varied considerably around this average estimate. Some studies included the main claim in the title, in which case coding could take seconds. Other studies required coders to consult several sections of the article to verify that a claim in question was indeed the main claim of the paper. In these cases coding could take several minutes. In every case, the main finding of the study was mentioned in the abstract of the article in which the study appeared.

With respect to identifying statistical information for each finding, however, we quickly realized that this would become challenging. By and large, main findings were associated with a number of different statistical results. Consider the following, example:

> In two experiments, we used a functional magnetic resonance (fMR)-repetition suppression paradigm to demonstrate that distinct frontal–parietal–temporal regions are sensitive to processing the scenarios or what participants imagined was happening in an event (e.g. medial prefrontal, posterior cingulate, temporal–parietal and middle temporal cortices are sensitive to

> the scenarios associated with future social events), people (medial prefrontal cortex), objects (inferior frontal and premotor cortices) and locations (posterior cingulate/retrosplenial, parahippocampal and posterior parietal cortices) that typically constitute simulations of personal future events. This pattern of results demonstrates that the neural substrates of these component features of event simulations can be reliably identified in the context of a task that requires participants to simulate complex, everyday future experiences. - Szpunar et al. (2014)

It is clear that many statistical results are being utilized in this statement, and it is not clear which, if any, would be more appropriate to serve as the results on which a replication value estimate is based. Many of the findings identified in our pilot had a similar structure to the example above. We suspect this finding structure will be common in the field of social fMRI, where hypotheses are often of the form "what does neural activity look like for task/manipulation/stimulus/group X?" and therefore relate to multiple aspects of the fMRI data collected. For the purposes of collecting statistical data for replication value estimation, it appears it would not be enough to simply identify the main finding of each study in our dataset. We would also have to determine, for each finding, which empirical results to extract statistical information from and how to resolve the common case where a finding is related to multiple statistical results. Due to the labor intensity this implies, we determined not to proceed with the coding of main findings in this project.

References

Szpunar, Karl K., Peggy L. St Jacques, Clifford A. Robbins, Gagan S. Wig, and Daniel L. Schacter. 2014. "Repetition-Related Reductions in Neural Activity Reveal Component Processes of Mental Simulation." *Social Cognitive and Affective Neuroscience* 9 (5): 712–22. https://doi.org/10.1093/scan/nst035.