1 Consistent activation differences versus differences
2 in consistent activation: Evaluating meta-analytic
3 contrasts

4 Vincent Küppers[1,2,✉], Edna C. Cieslik[3,1], Lennart Frahm[4,1], Felix Hoffstaedter[1,3], Simon
5 B. Eickhoff[3,1], Robert Langner[3,1], and Veronika I. Müller[1,3,✉]

6 [1] Institute of Neuroscience and Medicine, Brain and Behaviour (INM-7), Research
7 Center Jülich, Jülich, Germany

8 [2] Department of Nuclear Medicine, University Hospital and Medical Faculty, University
9 of Cologne, Cologne, Germany

10 [3] Institute of Systems Neuroscience, Medical Faculty and University Hospital
11 Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

12 [4] Department of Psychiatry, Psychotherapy and Psychosomatics, School of Medicine,
13 RWTH Aachen University, Aachen

14 ✉ Correspondence: Vincent Küppers <v.kueppers@fz-juelich.de>, Veronika I. Müller
15 <v.mueller@fz-juelich.de>

16 1   Abstract

17 Meta-analytic contrasts are a promising aspect of coordinate-based meta-analyses in
18 neuroimaging research as they facilitate the statistical comparison of two meta-analytic
19 results. They have been used for a multitude of comparisons, such as task conditions,
20 cognitive processes, and groups. However, it remains to be tested how the results of
21 meta-analytic contrasts relate to those of classic meta-analyses and vice versa. Here
22 we present a comprehensive empirical investigation of this issue using four datasets
23 from three different cognitive domains: working memory, working memory load,
24 cognitive interference processing, and emotional face processing. For all four datasets,

25 we compared the results of a standard meta-analysis across prototypical contrasts

26 (condition A > condition B) reported in individual experiments with those of a contrast

27 between two individual meta-analyses of the same conditions (meta-analysis condition

28 A > meta-analysis condition B). In the meta-analytic contrasts similar brain regions as in

29 the standard meta-analysis were found but with relatively distinct spatial activation

30 patterns. Additionally, fewer regions were revealed in the meta-analytic contrasts,

31 especially in areas where the conditions spatially overlapped. This can be ascribed to

32 the loss of information on the strength of activations in meta-analytic contrasts, across

33 which standard meta-analysis summarize. In one dataset, additional regions were found

34 in the meta-analytic contrast, potentially due to task effects. Our results demonstrate

35 that meta-analytic contrasts can yield similar results to standard meta-analyses but are

36 sparser. This confirms the overall validity, but also limited ability to capture all regions

37 found in standard meta-analyses. Notable differences observed in some cases indicate

38 that such contrasts cannot be taken as an easy substitute for classic meta-analyses of

39 experiment-level contrasts, warranting further research into the boundary conditions for

40 agreement.

41 **Keywords**

42 activation likelihood estimation, neuroimaging meta-analysis, method evaluation,

43 replicability, research synthesis, robust comparisons

2   Introduction

45 The neuroimaging literature of the last three decades has provided a wealth of findings

46 on structural and functional brain–behaviour relationships as well as brain-related

47 alterations in certain diseases. Meta-analytic approaches to neuroimaging results, such

48 as the widely used Activation Likelihood Estimation (ALE) algorithm (Eickhoff et al.,

49 2012; Turkeltaub et al., 2002), are important tools to consolidate these findings and to

50 overcome problems of individual studies (Müller, Cieslik, et al., 2018). Beyond testing

51 for brain regions consistently found across studies, meta-analyses also provide the

52 possibility to directly and statistically compare the results of two individual meta-

2

53 analyses, using conjunctions and meta-analytic contrasts (Eickhoff et al., 2011; Laird et
54 al., 2005). While conjunctions assess overlap in convergence, contrast analyses on the
55 meta-analytic level reveal those voxels of the brain in which convergence in one meta-
56 analysis (e.g. across studies on face processing) is significantly stronger than in another
57 (e.g., across studies on working memory). Meta-analytic contrasts therefore provide a
58 possibility to statistically compare aggregated neuroimaging data and, since their
59 introduction, have enjoyed great popularity for comparing different experimental
60 conditions (Owen et al., 2005; Swick et al., 2011; Wesley & Bickel, 2014) , mental
61 processes (Caspers et al., 2010; Gan et al., 2022; Langner et al., 2018) , or different
62 groups (Costa et al., 2021; Fehlbaum et al., 2021; Hill et al., 2014; Stevens & Hamann,
63 2012) .

64 Meta-analytic contrasts are particularly useful to examine previously untested (or poorly
65 studied) differences. Kogler et al. (2015) for example reported similarities and
66 differences between psychological and physiological stress that has not been tested
67 before in single fMRI studies and mainly found, contrary to previous assumptions,
68 differential convergence for both types of stress. Rottschy et al. (2012) assessed
69 differences in working memory based on stimulus material, task load and paradigms
70 and reported regional specific convergence within the left dorsolateral prefrontal cortex.
71 Importantly, meta-analytic contrasts can also help to break down larger cognitive
72 concepts into their subcomponents. For example, Kogler et al. (2020) found supporting
73 evidence for a multidimensional concept of empathy, Zhang et al. (2021) for the three
74 subcomponents of executive functions, and Langner et al. (2018) for differences
75 between top-down emotion and action regulation, which they confirmed through
76 additional analyses.

77 In addition, meta-analytic contrasts are also applied to test (previously not investigated)
78 group differences, such as different ages (Fehlbaum et al., 2021; Yaple et al., 2019;
79 Zhang et al., 2021) or clinical groups (Costa et al., 2021; Klugah-Brown et al., 2021).
80 This is an important potential of contrasts on the meta-analytic level as studies

81 comparing different clinical groups to each other as well as to a control group are quite

82 costly.

83 These examples illustrate some of the many uses of meta-analytic comparisons.

84 Especially, as the number of neuroimaging results continues to increase, many more

85 applications of meta-analytic comparisons become possible. While all of them can

86 theoretically also be carried out in individual fMRI studies, meta-analytic contrasts take

87 advantage of the wealth of already conducted experiments by looking at them from a

88 different angle and are a promising exploratory method for hypothesis generation.

89 However, while meta-analytic contrasts offer new possibilities for investigating different

90 concepts, it should be noted that the results of meta-analytic contrasts may not

91 necessarily reflect differences in brain activation, despite often being interpreted as

92 such. Classic ALE meta-analyses are conducted across coordinates derived from

93 individual neuroimaging studies, which typically reflect differences in activation strength

94 between two different brain states or participant groups. The results thus reflect the

95 convergence (i.e., consistency) of brain activation differences reported across these

96 studies (Eickhoff et al., 2012). ALE meta-analytic contrasts, on the other hand, compare

97 the results of two classic meta-analysis, testing for those voxels where convergence

98 (the ALE value) significantly differs between two meta-analysis. Meta-analytic contrast

99 analysis thus adds an additional layer on top of classic ALE analyses and is an

100 important tool for interpretation as it provides formal information if brain regions found in

101 one meta-analysis but not in the other, truly differ in their ALE scores.

102 It might be suggested that meta-analytic contrasts and classic meta-analyses of the

103 same comparison pick up similar mechanisms. However, as can be seen from the

104 description of the calculation above, meta-analytic contrasts differ conceptually. While

105 meta-analytic contrasts investigate differences in convergence without considering

106 activation strength, classic meta-analyses test for convergence across activation

107 strength. This distinction is further highlighted by two previous clinical meta-analyses

108 using both approaches to investigate changes in brain activation in autism (Costa et al.,

109 2021) and borderline personality disorder (Degasperi et al., 2021) compared with

110   healthy controls. Both studies found divergent results for the two approaches. Costa et

111   al. (2021) for example identified regions of consistent aberrant brain activation

112   computing a classic meta-analysis (across experiments of patients vs. controls).

113   However, a meta-analytic contrast of the same but conceptually different comparison

114   (meta-analysis across patients vs. meta-analysis across controls) did not. The authors

115   argue that differences in the magnitude of activation can be obscured when looking at

116   differences in convergence (i.e., meta-analytic contrasts). However, while highlighting

117   the importance to distinguish between the two approaches, previous results can only be

118   generalized to group comparisons. Importantly, clinical meta-analyses often consolidate

119   findings found for a specific patient group independent of a particular process and are

120   therefore quite heterogeneous in the tasks and processes investigated (Müller et al.,

121   2017), leading to overall less convergence for all analyses. Results reported by Costa et

122   al. (2021) and Degasperi et al. (2021) may therefore be affected by this heterogeneity

123   and may not be transferable to meta-analyses that investigate different task conditions.

124   Thus, the interpretation of meta-analytic contrast results and their ability to capture

125   effects observed at the experimental level are still not fully clarified. The increasing

126   popularity in computing these contrasts and the resulting interest in understanding the

127   findings necessitates a systemic investigation.

128   The current study aimed to provide deeper insights into the interpretability and validity of

129   meta-analytic contrasts by conducting a comprehensive empirical evaluation. We

130   compared the results of meta-analytic contrasts with the results of individual meta-

131   analyses across experimental comparisons. This approach is similar to Costa et al.

132   (2021) and Degasperi et al. (2021), who computed group comparisons (patients >

133   controls; meta-analysis patients > meta-analysis controls). However, in order to keep

134   the complexity low and to avoid additional confounding effects at the level of group

135   comparisons, we took a step back and focused on comparing prototypical task

136   conditions in three different cognitive domains. Thus, in each domain, we compared the

137   results of a meta-analysis across contrasts between two conditions (condition A >

138   condition B) with the results of a meta-analytic contrast that contrasted two meta-

139   analyses of the same two conditions with each other (meta-analysis condition A > meta-

140   analysis conditions B).

141   # 3   Methods

142   In this paper, the standard meta-analysis across experiment-level contrasts is

143   abbreviated to *CM* (contrast-meta); the meta-analytic contrast between two meta-

144   analyses is abbreviated to *MC* (meta-contrast). It is important to note that the input-data

145   (coordinates we obtained from the individual neuroimaging studies) required for the

146   computation of the two approaches is different. In the first case (CM), coordinates are

147   needed for the contrast between *condition A > condition B*, whereas in the second case

148   (MC), the main effect coordinates are needed for *condition A > baseline* and

149   coordinates for *condition B > baseline*. The availability of the coordinates determines

150   which method can be used.

151   To conduct a comprehensive empirical study, we collected four sets of data from

152   different cognitive domains. *Working memory 2-back > 0-back*: based on the n-back

153   task, the 2-back (condition A) was contrasted with the 0-back (condition B) condition. To

154   compute the contrast-meta, experiments reporting 2-back > 0-back (A>B) were

155   collected. For the meta-contrast, experiments reporting 2-back > baseline (A) and 0-

156   back > baseline (B) were separately collected, two individual meta-analyses (meta-

157   analysis across A and one across B) computed and the results statistically compared in

158   the meta-contrast. *Working memory 2-back > 1-back*: experiments reporting 2-back > 1-

159   back (A>B) for the CM and 2-back > 0-back (A) and 1-back > 0-back (B) for the MC

160   were collected. *Cognitive interference processing*: experiments of the colour-word

161   Stroop task were collected. Incongruent > congruent (A>B) experiments were collected

162   for the CM, whereas incongruent > control/baseline (A) and congruent >

163   control/baseline (B) experiments were gathered for the MC. *Emotional face processing*:

164   experiments across different tasks using emotional and neutral face stimuli were

165   collected. Experiments of the emotional > neutral faces (A>B) comparison were used for

166 the CM. Emotional faces > control/baseline (A) and neutral faces > control/baseline (B)

167 for meta-contrast (see Table 1).

168 *Table 1: Datasets and experiments (contrasts at experimental level) included in*

169 *contrast-meta (CM) and meta-contrast (MC)*

Domain (Dataset)	Experiments included in CM	Experiments included in MC - condition A	Experiments included in MC - condition B
Working memory (WM: 2-back > 0-back)	2-back > 0-back	2-back > baseline	0-back > baseline
Working memory load (WM: 2-back > 1-back)	2-back > 1-back	2-back > 0-back	1-back > 0-back
Cognitive interference processing (interference)	incongruent > congruent	incongruent > control; incongruent > baseline	congruent > control; congruent > baseline
Emotional face processing (emo)	emotional > neutral faces	emotional faces > control; emotional faces > baseline	neutral faces > control; neutral faces > baseline

3.1 Analysis approach

171 As described above, we used two different meta-analytic approaches for computing the

172 same contrast (e.g., *condition A > condition B*): First, an ALE meta-analysis was

173 calculated across experiments that delineated the contrast of interest on the

174 experimental level (CM). The CM thus includes experiments reporting results that

175 compare condition A with condition B (*condition A > condition B*). This type of meta-

176 analysis thus revealed spatial convergence of differences in brain activations between

177 conditions across experiments (Müller, Cieslik, et al., 2018) (compare Figure 1).

178 Second, the contrast of interest was calculated by contrast analysis between the results

179 of two meta-analyses (MC). To do so, we first calculated two separate meta-analyses,

180 one across experiments contrasting *condition A* with a control or baseline condition and

181 once across experiments contrasting *condition B* with a control or baseline condition.

182 Importantly, a significant convergence of a brain region in the meta-analysis across

183 *condition A* and a lack of convergence of the same brain region in the meta-analysis

184 across *condition B* does not necessarily indicate that *condition A* differs in convergence

from *condition B*. Thus, to formally test significant difference in convergence, the two meta-analyses were contrasted against each other using meta-analytic contrast analysis resulting in the same contrast of interest (*condition A > condition B*) as CM but calculated in a different way. In contrast to the CM, the MC results reflect differences in the across-study convergence of brain activation between conditions.
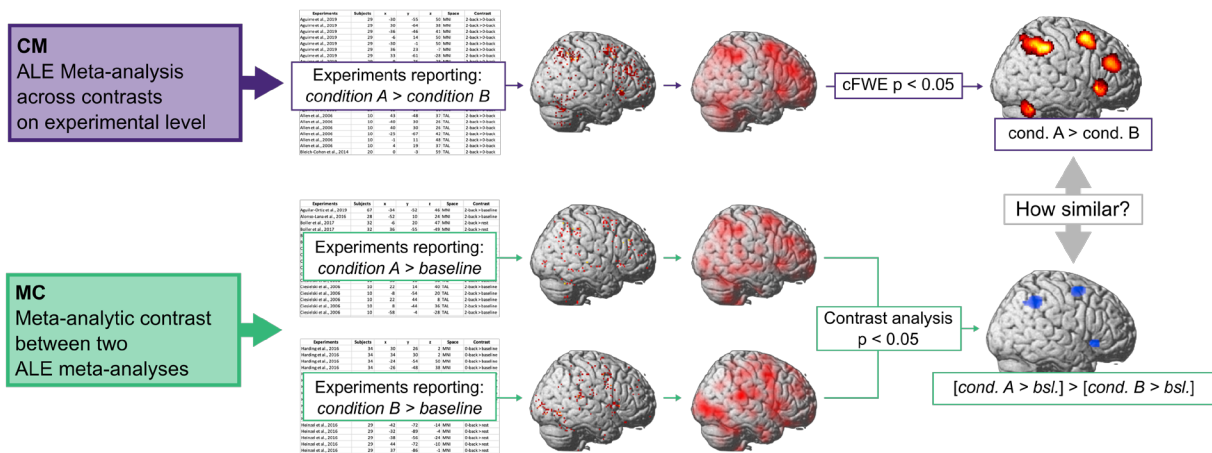

Figure 1: Schematic depiction of the analytic approach: Comparison of the results of a meta-analysis of experiments contrasting condition A with condition B (CM) with a meta-analytic contrast between a meta-analysis of experiments contrasting condition A with a baseline and another meta-analysis contrasting condition B with a baseline (MC).

3.1.1 Activation likelihood estimation

For each dataset, we conducted three meta-analyses: one for the meta-analysis across experimental contrasts (CM), and two for the meta-analytic contrasts (MC). All meta-analyses were computed based on the activation likelihood estimation algorithm (Eickhoff et al., 2009, 2012; Turkeltaub et al., 2002, 2012). Here we used an in-house Python implementation of the algorithm (https://github.com/LenFrahm/pyALE).

In ALE, the foci of all included experiments are treated as centers of Gaussian probability distributions which captures the spatial uncertainty of the coordinates. The width of these distributions is modeled based on empirical data of between-template and between-participant variance. The between-participant variance is weighted by the

205 number of participants of the respective study as a larger sample size is considered as

206 spatially more reliable and therefore modeled with a denser Gaussian than experiments

207 with a smaller sample size. For each experiment, a modeled activation map is

208 generated by aggregating the probabilities of all reported foci for each voxel. To account

209 for experiments reporting multiple foci close to each other, each voxel only receives the

210 largest possible probability value from all foci close to it (Turkeltaub et al., 2012). Then,

211 the probabilities are combined by taking the voxel-wise union across all modeled

212 activation maps of each individual experiment. The resulting map of voxel-wise ALE

213 scores thus reflects the convergence of results across experiments in each voxel of the

214 brain (Eickhoff et al., 2009).

215 Next, the ALE scores are tested against a null distribution of random spatial

216 associations to distinguish true from random convergence and results are thresholded

217 at p<0.001. The null distribution is calculated using an analytic procedure based on a

218 nonlinear histogram algorithm (Eickhoff et al., 2012).

219 Finally, a permutation approach is used to correct for multiple comparisons (cFWE p <

220 0.05) by comparing the cluster-sizes to a null distribution of cluster sizes. This null

221 distribution is created by randomly distributing foci (except location all other properties

222 of the foci are held constant) within a grey-matter mask, calculating an ALE analysis in

223 the same manner as with the real data and recording the maximum cluster size found in

224 this analysis. This procedure is repeated 10,000 times resulting in an empirical null

225 distribution of cluster sizes. A cluster is considered significant if its size exceeds the

226 sizes of 95% of those random permutations.

227 3.1.2 ALE meta-analytic contrasts

228 To calculate MCs for each dataset, we contrasted the two ALE meta-analyses across

229 the coordinates of the main effects (*condition A > baseline* and *condition B > baseline*).

230 The ALE contrast analysis is a statistical comparison between the results of two meta-

231 analyses (Eickhoff et al., 2011). Here we used an in-house implementation in Python,

232 similar to the current implementation in BrainMap, GingerALE (Version 3.0.2). Meta-

analytic contrast analyses were performed by first calculating two separate meta-analyses and calculating the voxel-wise difference score. These difference scores are then compared to an empirical null distribution of ALE difference scores under the assumption of exchangeability. This was done by pooling the experiments of both conditions, shuffling them and randomly splitting them into two groups of the original size of experiments. ALE difference scores between these random groups are then recorded for every voxel. This process is repeated 10000-times and the real differences in ALE scores are tested against this null distribution of difference scores. Here, a threshold of $p > 0.95$ was used (i.e., the observed probability in the difference between ALE values per voxel needed to be equal to or higher than 95% chance level), the results were inclusively masked by the respective main effect of the condition of interest, and an additional cluster extent threshold of $k = 5$ was applied.

3.2 Data collection and datasets

We conducted a comprehensive literature search to construct, as mentioned above, four datasets for three distinct cognitive domains: working memory (WM), cognitive interference processing (interference), and emotional face processing (emo). These domains were selected based on their extensive investigation in the neuroimaging literature and the availability of sufficient experiments to calculate both a meta-analysis across experimental contrasts (CM) as well as a meta-analytic contrast between two meta-analyses (MC). We did not seek to obtain an ethics vote for this study as our analyses were based solely on previously published aggregated data and did not include individual participant data.

The datasets were constructed from previous meta-analyses (Cieslik et al., 2015; Langner & Eickhoff, 2013; Müller, Höhner, et al., 2018; Rottschy et al., 2012) and extended by tracing references from additional meta-analyses and conducting a comprehensive literature search using the "PubMed" (https://pubmed.ncbi.nlm.nih.gov/) and "Web of Science" (http://webofknowledge.com/) search engines. Different variations of the keywords "fMRI" and "PET" were combined with condition- and domain-specific

261 search terms. We included studies, that reported results of at least one of the contrasts

262 of interest (A>B, A or B) i.e., reporting of all three contrasts of interest was not a

263 requirement for inclusion since only few studies reported all contrasts. In accordance

264 with the general guidelines (Müller, Cieslik, et al., 2018), only whole brain comparisons

265 (scan and analysis) from healthy adults (mean age 18 or older) that reported results in

266 standard anatomical space (TAL or MNI) were included. We excluded intervention,

267 treatment, or induction studies (e.g., medication, TMS). However, if a study reported

268 results of a baseline measurement before an intervention, the corresponding contrast

269 was included. In cases where a paper reported multiple experiments for the same

270 condition of interest in the same group of participants, the coordinates were pooled and

271 treated as one experiment. Results reported in one paper but obtained from different

272 participant groups were treated as separate experiments. Additionally, we reached out

273 to authors (Agostini et al., 2017; Aguirre et al., 2019; Campanella et al., 2013; Cullen et

274 al., 2016; Daamen et al., 2015; Dan et al., 2019; Fukuda et al., 2019; Ghavidel et al.,

275 2020; Habel et al., 2007; Harding et al., 2016; Jung et al., 2018; Kaminski et al., 2020;

276 King et al., 2015; Köhler et al., 2016; Kowalczyk et al., 2021; Kozasa et al., 2018;

277 Kronbichler et al., 2018; Krug et al., 2008; Lahr et al., 2018; Li et al., 2019; Luethi et al.,

278 2016; Miró-Padilla et al., 2019; Papalini et al., 2019; Peven et al., 2019; Schlagenhauf

279 et al., 2008; Schmidt et al., 2012; Schneider et al., 2007; Shashidhara et al., 2020;

280 Smits et al., 2009; Szabó et al., 2019; Takeuchi et al., 2018; Vacchi et al., 2017; van der

281 Horn et al., 2016; Wagner et al., 2015; Wallentin et al., 2015; Wu et al., 2017; Ye &

282 Zhou, 2009) via email to request pertinent additional results (see Figure 2).

283 Literature search and coding was performed by one author and the resulting datasets

284 were checked for eligibility and correctness by a second author. Literature search was

285 performed up to 2021-12-09.

286 Details about the experimental designs, conditions, and specific tasks for each dataset

287 can be found in the supplementary material.

288 In the working memory (2-back > 0-back and 2-back > 1-back) and interference

289 processing datasets, the experiments showed a relatively high degree of homogeneity.

290   Specifically, within the working memory domain, all experiments used the n-back task,

291   with variations between experiments only in the stimuli used (letters, numbers, pictures

292   etc). Similarly, in the interference processing domain, experiments used the colour-word

293   Stroop task, with only some variations in the stimuli (different colours, number of

294   colours). Additionally, the contrasting condition (> *baseline*) of the experiments used for

295   the calculation of the MC remained relatively consistent in the working memory dataset.

296   However, in the interference processing dataset, there was some variation, with rest

297   and control (e.g., words, symbols) as contrasting conditions.

298   In contrast, the emotional face processing dataset showed greater heterogeneity

299   compared to the other two domains. This dataset encompassed experiments involving

300   different tasks (such as gender-discrimination, passive viewing, and emotion-matching)

301   along with a wide range of stimuli (different face datasets, different emotions etc.). The

302   contrasting conditions in this dataset were more diverse, including rest and control

303   conditions with varying characteristics such as shapes, objects, scenes, and others.

304   Further details about the four datasets can be found in the supplementary material

305   (Table S4-6), providing a comprehensive description of the experimental designs and

306   conditions.

307   Figure 2 provides a visual representation of the literature search process, including the

308   number of authors contacted and the contrasts of interest.
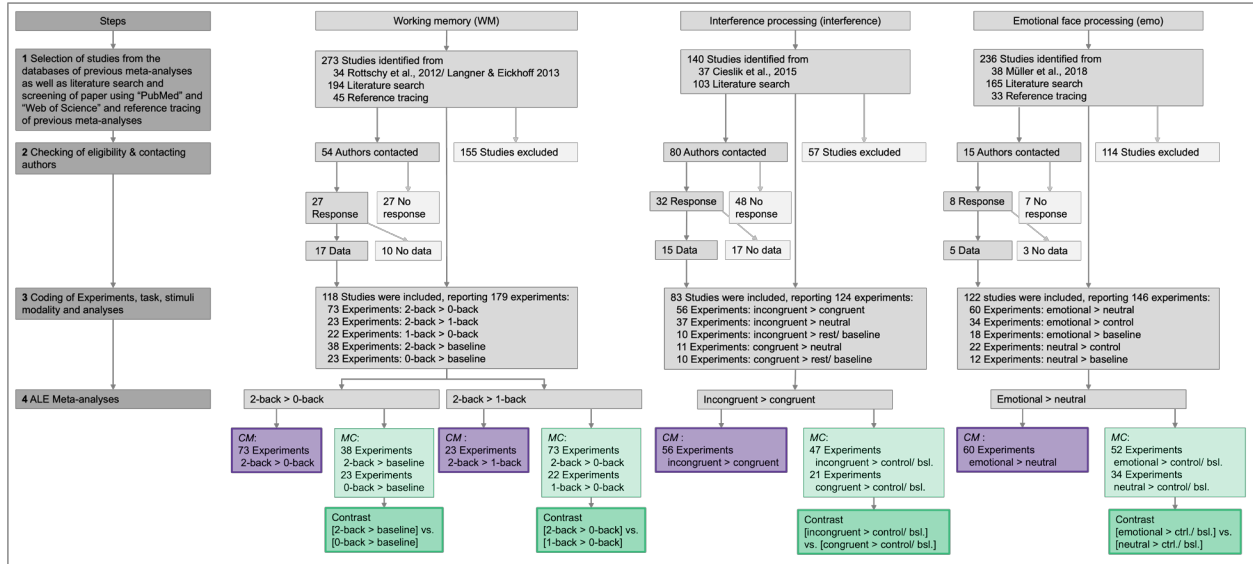
309

Figure 2: Flowchart depicting the creation of four datasets from three cognitive domains and the subsequent analyses.

3.2.1 Comparison of meta-contrast and contrast-meta results

To quantify the extent to which the MC results were similar to the CM results, different metrics were applied for voxel-, cluster-, and peak-related comparisons.

Voxel-wise comparison

First, we evaluated similarity based on a whole-brain voxel-wise comparison of the two result maps. First, we binarized both maps, i.e., treated significant voxels as 1's and non-significant voxels as 0's and calculated the Jaccard similarity coefficient (Jaccard, 1901; Maitra, 2010) as well as the sensitivity and precision. Jaccard coefficient was calculated as the intersection of all significant voxels between CM and MC divided by the union of significant voxels (Equation (1)).

$$Jaccard\ coefficient = \frac{CM \cap MC}{CM \cup MC} \qquad (1)$$

Sensitivity was assessed by calculating the intersection of all significant voxels between CM and MC divided by all significant voxels of CM (Equation (2)). This metric reflects how well the MC reveals voxels that are also significant in the CM analysis.

$$326 \qquad Sensitivity = \frac{CM \cap MC}{CM} \qquad (2)$$

327 Precision in contrast shows the proportion of all significant MC voxels that lie within the
328 CM network (Equation (3)).

$$329 \qquad Precision = \frac{CM \cap MC}{MC} \qquad (3)$$

330 **Cluster-wise comparison**

331 As a second measure of similarity, a comparison of clusters was chosen, providing a
332 similarity measure that is relatively independent of the size of clusters and more based
333 on the spatial location of results. Here we assess how many clusters of one map
334 overlap with clusters from the other map. Clusters are counted as overlapping if at least
335 one voxel of both maps is overlapping (Bossier et al., 2020). Cluster-wise sensitivity
336 was computed by first dividing the number of CM clusters overlapping with clusters from
337 MC by the total number of CM clusters. Precision was assessed by first dividing the
338 number of MC clusters overlapping with clusters from CM by the total number of MC
339 clusters.

340 **Peak-wise comparison**

341 As the cluster-wise comparison does not take into account cases where the same
342 anatomical region is revealed in both analyses without direct overlap, e.g. in the case of
343 relatively small clusters, similarity between meta-analytic maps was additionally
344 assessed by assessing the location of local maxima. This was done by extracting peaks
345 from CM and MC using the FSL cluster command on the uncorrected thresholded z-
346 score maps. We extracted all local maxima with a minimum distance of 8 mm (default
347 distance in SPM12 - Statistical Parametric Mapping software) and evaluated the
348 proximity to the nearest peak in the opposite map. In particular, for each local maxima
349 from CM we calculated the Euclidean distance to the next peak from MC. Likewise, for
350 all local maxima from MC we determined the closest peak from CM. Median distances
351 in each direction (CM to MC and MC to CM) were computed and reported in millimeters.

352 It should be noted that the comparison was conducted using the significant result maps

353 rather than the unthresholded maps. This approach allowed us to assess the similarity

354 of the results with those that would typically be reported in a meta-analytic study.

3.2.2 Comparison of MC to a contrast derived from large-sample single study

357 We also compared MC of 2-back > 0-back to the results of a contrast derived from a

358 highly powered large single study dataset, calculating voxel-wise, cluster-wise, and

359 peak-wise comparisons in the same way as for comparisons with CM. For this, we used

360 task-based fMRI data from the working memory task of an unrelated sample of the

361 Human Connectome Project (HCP) (Van Essen et al., 2013). Details regarding the

362 dataset, sample and preprocessing as well as the results can be found in the

363 supplementary material. The maps of the two "comparison maps," i.e., CM and HCP

364 demonstrated a voxel-wise Jaccard index of 0.27.

4   Results

4.1   Description of the datasets

367 To investigate the effects captured in meta-analytic contrasts (MCs) in comparison with

368 meta-analyses across the same contrasts at the experimental level (CMs), we

369 compared four MC analyses and corresponding CM analyses across three different

370 cognitive domains. The final datasets included 134 experiments for working memory: 2-

371 back > 0-back, 118 experiments for working memory: 2-back > 1-back, 124 experiments

372 for interference processing, and 146 experiments for emotional face processing. Most

373 studies reported experiments for only one condition of interest (A, B or A>B), but there

374 were also that provided experiments for two or all three contrasts of interest. All three

375 contrasts of interest (i.e., A, B, A>B) were available for 12 studies in WM: 2-back > 0-

376 back, for 6 studies in WM: 2-back > 1-back, for 15 studies in interference processing,

377 and for only 3 studies in emo dataset. Both sets of coordinates required to compute the

378 MC (i.e., A, B) were available for 16 studies in WM: 2-back > 0-back, 17 studies in WM:

379 2-back > 1-back, 20 studies in interference, and 17 studies in emo datasets.

4.2   Comparison between CM and MC

381 Overall, voxel-wise comparisons revealed disparities in the number of significant voxels

382 between CM and MC, with CM exhibiting a larger network for all datasets (see Figure 3

383 – 6). Thus, CM is in general more sensitive than MC. For the working memory (2-back >

384 0-back) and interference processing contrasts, the voxels identified by MC largely

385 aligned with the CM network. However, this correspondence was not observed for the

386 2-back > 1-back and emotional face processing contrasts. At the cluster level, several of

387 the MC clusters overlapped with those of CM.

4.2.1 Working memory

389 **2-back > 0-back**. As shown in Figure 3, the classic CM yielded a more extensive

390 network in terms of regions and significant voxels, as compared to the MC results.

391 Voxel-wise Jaccard coefficients and sensitivity are thus rather low (~0.1, see

392 supplementary Table S1). However, all regions showing significantly stronger

393 convergence in the 2-back > baseline meta-analysis, as compared to the 0-back >

394 baseline meta-analysis (MC), were located entirely within the CM network, reflected by

395 high precision (0.98) and perfect cluster and peak overlaps (Figure 3). Regions

396 delineated in both types of contrast analyses included bilateral intraparietal sulcus (IPS),

397 dorsal premotor cortex (dPMC), (pre-)supplementary motor area (SMA), anterior insula

398 and left cerebellum; in contrast, 3 out of 4 cerebellar regions, left basal ganglia, bilateral

399 dorsolateral prefrontal cortex (dlPFC), frontal pole as well as inferior frontal junction

400 (IFJ) were not identified in MC, which is reflected in low cluster-wise sensitivity (0.47)

401 and relatively far median peak distance (14.4 mm).

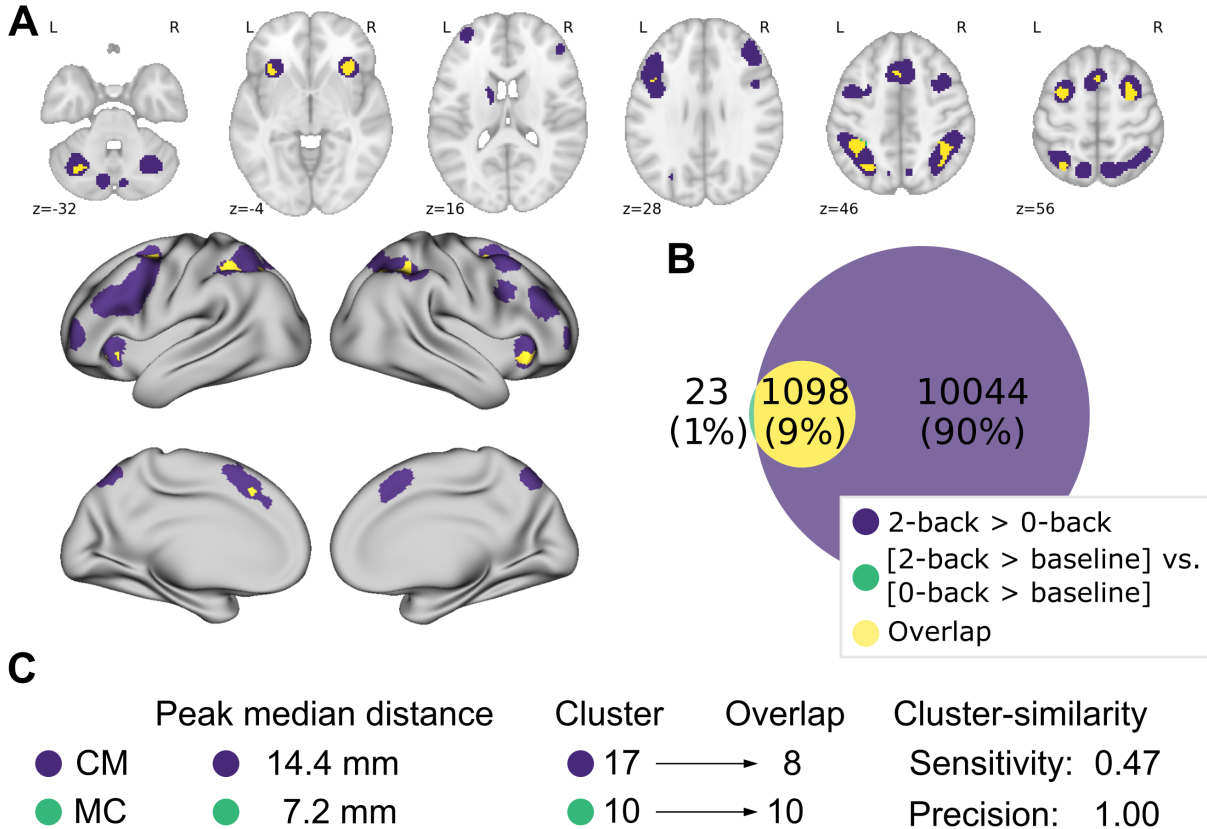**Working memory: 2-back > 0-back**

Figure 3: Comparison of meta-analytic contrast (MC, green) and contrast-meta (CM, purple) for 2-back versus 0-back. A, MC and CM projections on MNI152 volume and fsLR (FreeSurfer surface template) surface. B, Venn diagram depicting the absolute voxel-wise overlap between the two maps (yellow). C, median distance between CM peaks and their nearest MC peaks, and vice versa. Cluster-wise overlap by count of CM and MC clusters overlapping with any cluster of the opposite map. Sensitivity, shows proportion of overlapping CM clusters with MC clusters, precision as overlapping MC clusters with CM.

2-back > 1-back. In a second dataset based on the WM domain, we focused on conditions with different working memory loads (i.e., 2-back > 1-back). At the voxel level, both networks seem to differ substantially (compare Figure 4 and supplementary Table S1). However, many clusters of MC and CM were located directly next to each other and therefore the comparison by cluster and peak locations revealed relatively

17

416 good correspondence of MC and CM (compare Figure 4). Regions for which significant

417 differences were found in both contrast analyses included the bilateral IPS, dlPFC, (pre-

418)SMA, and 3 clusters in the cerebellum. Left dorsal premotor cortex and right anterior

419 insula was only identified in CM while bilateral frontal pole clusters and left insula was
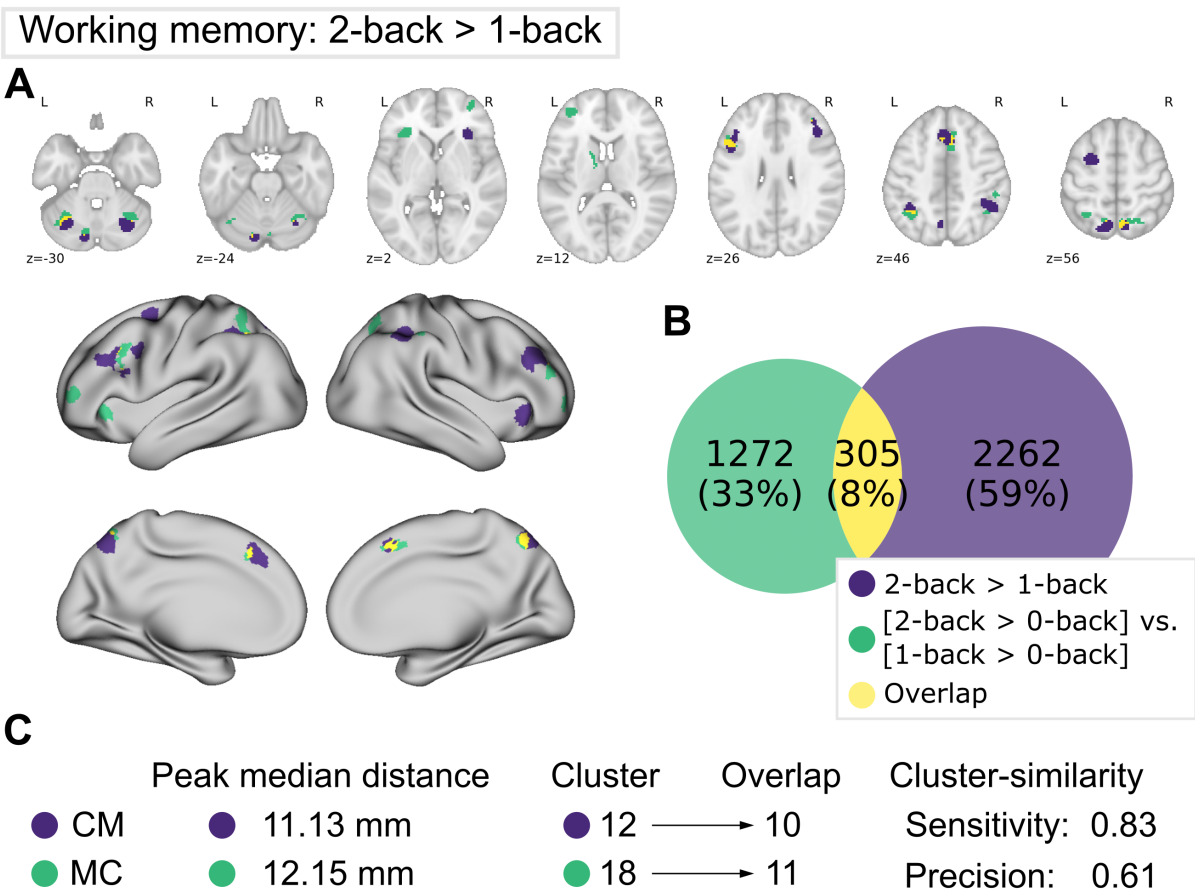
420 only found for MC.

Working memory: 2-back > 1-back

A

| z=-30 | z=-24 | z=2 | z=12 | z=26 | z=46 | z=56 |

B

1272 (33%) 305 (8%) 2262 (59%)

● 2-back > 1-back
● [2-back > 0-back] vs. [1-back > 0-back]
● Overlap

C

		Peak median distance		Cluster	Overlap	Cluster-similarity
● CM		● 11.13 mm		● 12 → 10		Sensitivity: 0.83
● MC		● 12.15 mm		● 18 → 11		Precision: 0.61

421

422 *Figure 4: Comparison of meta-analytic contrast (MC, green) and contrast-meta (CM,*

423 *purple) for the working memory 2-back versus 1-back dataset. A, MC and CM*

424 *projections on MNI152 volume and fsLR surface. B, Venn diagram depicting voxel-wise*

425 *overlap between CM and MC (yellow). C, proximities of peaks and cluster-wise overlap*

426 *between the maps. Refer to previous caption for detailed descriptions.*

4.2.2 Interference processing

The comparison of the interference meta-analytic contrast (MC) with the results of the contrast-meta (CM) yielded a similar pattern as seen for the WM 2-back > 0-back comparison (see Figure 5). However, for the interference contrast, higher voxel-wise similarity was found together with lower precision (supplementary Table S1). The results of the peak comparison showed a moderate correspondence from MC to CM (with a median distance of 10 mm), while this was relatively lower for CM to MC (with a median distance of 21.3 mm). Similarly, the cluster-wise comparison showed that almost all regions revealed by MC have overlapping clusters in CM, while many clusters revealed by CM had no correspondence in MC (compare Figure 5). Regions found in both contrast approaches comprised bilateral anterior insula, (pre-)SMA, midcingulate cortex, left IPS, and left dlPFC. For CM bilateral thalamus, right cerebellum and left fusiform gyrus was additionally identified, while MC revealed a cluster in right IPS.

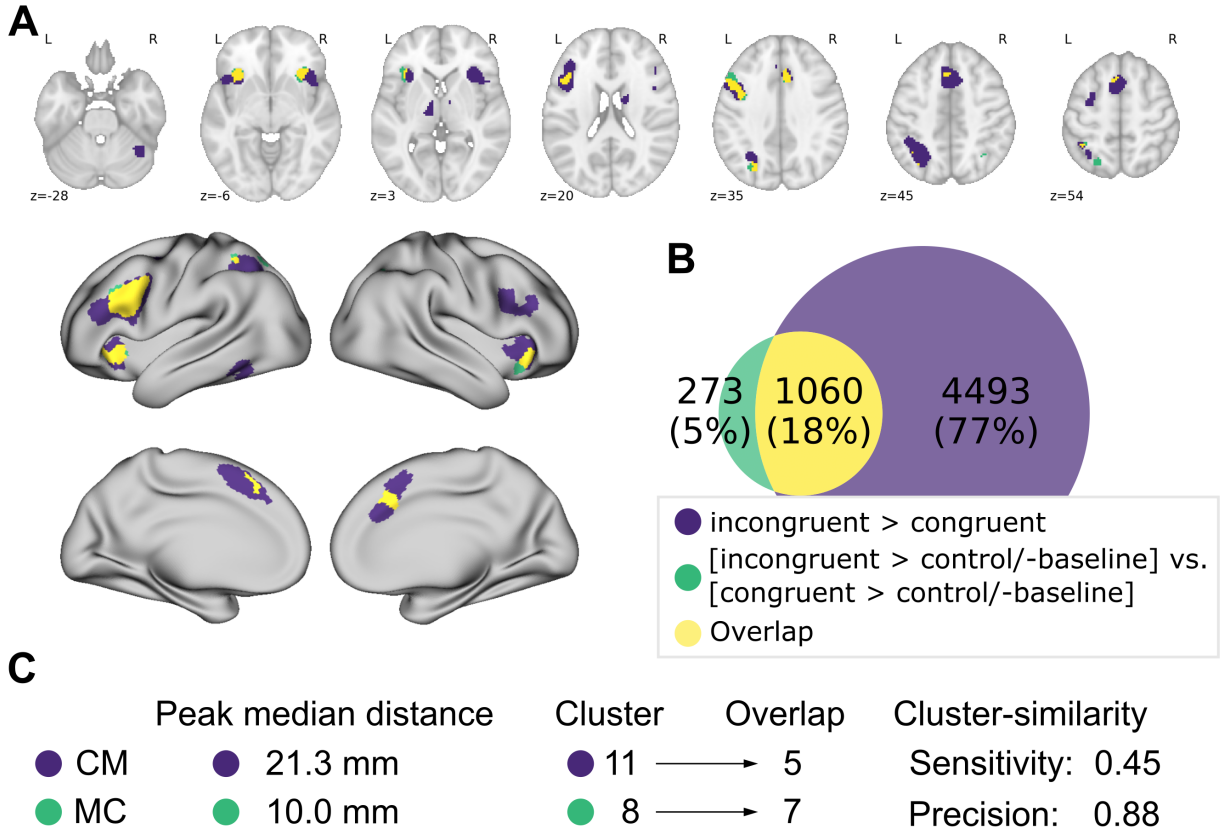Interference processing: incongruent > congruent

A

z=-28 z=-6 z=3 z=20 z=35 z=45 z=54

B

273 (5%)   1060 (18%)   4493 (77%)

- incongruent > congruent
- [incongruent > control/-baseline] vs. [congruent > control/-baseline]
- Overlap

C

	Peak median distance	Cluster	Overlap	Cluster-similarity
CM	21.3 mm	11 ⟶ 5		Sensitivity: 0.45
MC	10.0 mm	8 ⟶ 7		Precision: 0.88

Figure 5: Comparison of meta-analytic contrast (MC, green) and contrast-meta (CM, purple) for the interference dataset. A, MC and CM projections on MNI152 volume and fsLR surface. B, Venn diagram depicting voxel-wise overlap between CM and MC (yellow). C, proximities of peaks and cluster-wise overlap between the maps. Refer to previous caption for detailed descriptions.

4.2.3 Emotional face processing

For emotional face processing, CM and MC revealed networks with little to no similarity (see Figure 6). The only regional correspondence was observed in the left lateral occipital cortex, fusiform gyrus, and left amygdala. In contrast to the other domains (WM: 2-back > 0-back and interference), we observed twice as many clusters and peaks for MC than for CM (see Figure 6). In both analyses bilateral lateral occipital gyrus, left amygdala and left fusiform gyrus were found (in slightly different locations).

20

453 For CM right fusiform gyrus and left lateral orbitofrontal cortex was additionally found,

454 whereas MC identified additional clusters in bilateral dlPFC and IPS, left hippocampus,

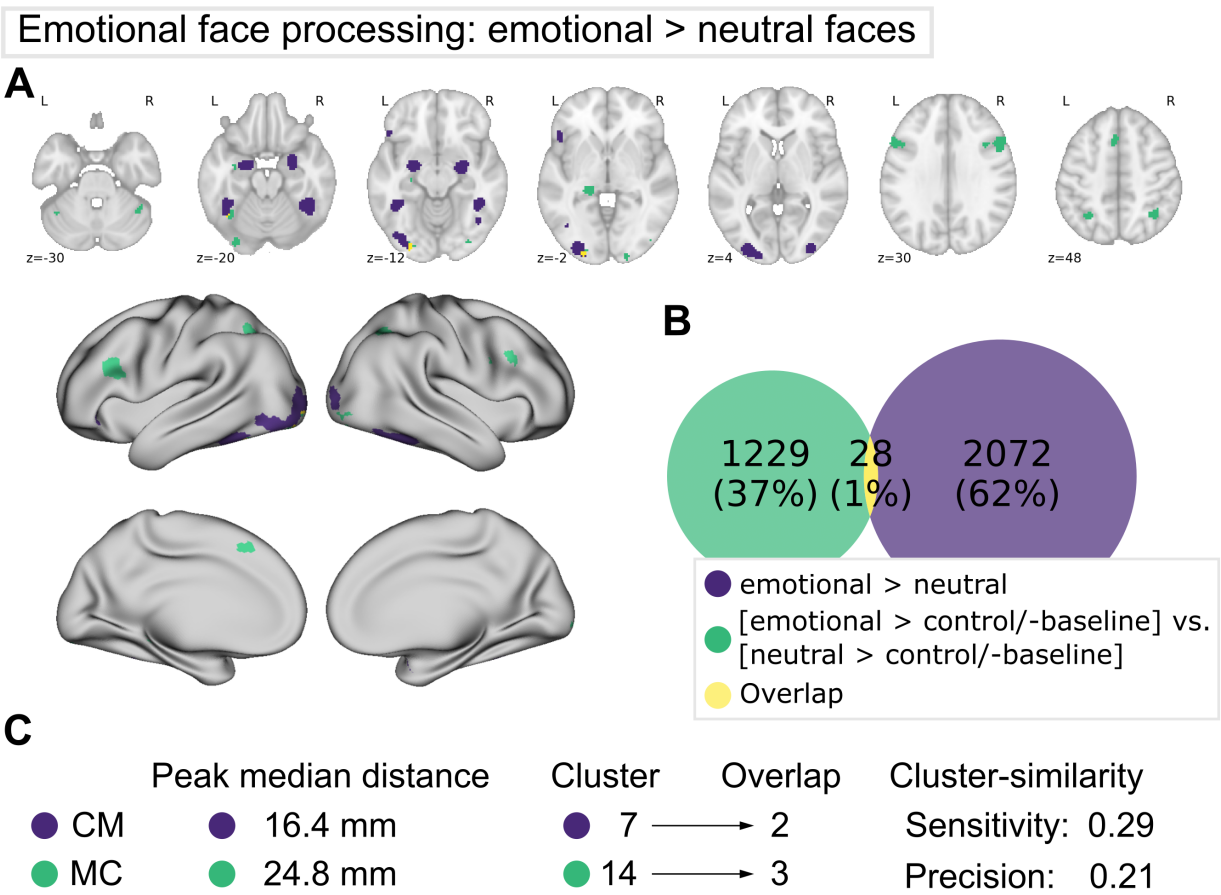455 right cerebellum, bilateral occipital pole as well as pre-SMA.

Emotional face processing: emotional > neutral faces

Figure 6: Comparison of meta-analytic contrast (MC, green) and contrast-meta (CM, purple) for the emotional face processing dataset. A, MC and CM projections on MNI152 volume and fsLR surface. B, Venn diagram depicting voxel-wise overlap between CM and MC (yellow). C, proximities of peaks and cluster-wise overlap between the maps. Refer to previous caption for detailed descriptions.

5   Discussion

463 Meta-analytic studies often test for differences between various mental faculties,

464 groups, and other experimental factors that are not sufficiently enough contrasted in the

465   neuroimaging literature (Fehlbaum et al., 2021; Klugah-Brown et al., 2021; Swick et al.,

466   2011) by computing separate meta-analyses for each condition of interest. However,

467   the results of individual meta-analyses do not provide any information about where in

468   the brain convergence differs between these analyses. For instance, the presence of

469   convergence in a specific brain region found in one meta-analysis and absence of the

470   same region in another one does not necessarily imply that one of them shows less

471   consistency across experiments than the other. Meta-analytic contrast analyses are

472   thus important tools to formally test for these differences and should be provided in any

473   meta-analytic study that claims to interpret differences between individual analyses

474   results. However, while the statistical relevance and usefulness of meta-analytic

475   contrast is out of question, it is not fully evaluated which exact conclusions can be

476   drawn from the results. This empirical investigation aimed to assess the extent to which

477   meta-analytic contrasts (MC), as implemented in the ALE meta-analysis framework,

478   reflect the effects observed in standard meta-analyses across experiment-level

479   contrasts (CM). Results revealed that meta-analytic contrasts revealed in general less

480   differences between conditions than CM together with a high rate of precision for most

481   datasets, except emotion processing, the dataset where tasks, control conditions and

482   stimuli varied most. Therefore, for most datasets, regions found in meta-analytic

483   contrast analysis can quite confidently be interpreted in a similar way as results of

484   classic CMs, i.e., consistent activation differences between conditions. However, an

485   absence of differences in meta-analytic contrast results does not necessarily imply the

486   absence of consistent activation differences. This is especially true for regions that

487   overlap across the two meta-analyses that are compared to each other, as well as to

488   low-powered and/or datasets with a lot of experimental and methodological variation.

489   Thus, meta-analytic contrasts, in addition to providing statistical evidence of differences

490   in convergence between the results of separate analyses, are well suited for

491   complementary and exploratory meta-analytic investigations, especially in situations

492   where there is insufficient literature reporting the exact contrast of interest. However, the

493   results of MC should always be interpreted considering the specific characteristics of

494 the datasets and potential systematic confounds be assessed that could affect the
495 results.

5.1 Functional relevance of identified regions

497 When looking at the specific regions revealed for the working memory and interference
498 processing domains, CM and MC both reveal networks including, frontal and parietal
499 regions as well as the anterior insula and (pre)SMA. These regions are part of the
500 multiple demand (Camilleri et al., 2018; Duncan, 2010), central executive (Menon,
501 2011) or cognitive control (Cole & Schneider, 2007) networks and are thus scientifically
502 meaningful to be involved in working memory and interference processing. Interestingly,
503 most regions that are additionally found in CM or MC, like the more anterior dlPFC,
504 dPMC and subcortical regions (basal ganglia and thalamus) are part of the so called
505 extended multiple demand network (Camilleri et al., 2018), regions involved in executive
506 functions but, in contrast to the core multiple demand regions, more dependent on
507 specific cognitive demands. These regions might therefore show smaller and less
508 robust effects and are potentially only found when contrasted against a specific high-
509 level control condition or in experiments with specific task characteristics.

510 For the emotion processing domain, in turn, there were only few regions that were found
511 in both contrast approaches. While CM primarily identified classical regions of (facial)
512 emotion processing, i.e. bilateral amygdala, fusiform and inferior occipital gyrus as well
513 as the left lateral orbitofrontal cortex (Adolphs, 2002; Dolcos et al., 2011), the contrast at
514 the meta-analytic level (MC) revealed many regions (like dlPFC, IPS, preSMA,
515 hippocampus) that are more implicated in emotional control as well as increased task
516 demand (Dolcos et al., 2011; Duncan, 2010; Phillips et al., 2003). Thus, MC seems to
517 pick up cognitive processes during emotion perception to a stronger degree than CM.

23

5.2 Regions showing convergence of differences but no differences in convergence

The results of this study revealed relatively low voxel-level similarity between MC and CM. This can be attributed to the generally lower number of significant voxels in MC. When looking at the results from a regional perspective, 45-83 % of the CM result clusters were also obtained in MC for WM and interference processing. However, still about half of the regions remained undetected and much less in the dataset of emotional processing. This is consistent with the results of Costa et al. (2021), reporting less differences between groups when using MC, relative to contrast meta-analyses.

Power. The fact that less voxels/regions show convergence in CM but not in MC might be due to low power in the individual meta-analyses, with more experiments needed for MC to detect the same effects as CM. In general, all the individual meta-analyses include enough experiments, i.e., n > 21 experiments, for detecting strong effects and most of them are decently powered for detecting medium-size effects (Eickhoff et al., 2016). However, more power might be needed for finding differences in convergence (via MCs). This is supported by the results of the 2-back > 1-back comparison, exhibiting the highest cluster-level sensitivity and lowest median peak distance (CM to MC peaks) with 73 experiments included in the meta-analysis of 2-back > 0-back. When reducing the number of experiments to 22 experiments in the deterministically matched analyses (see supplementary Figure S1), sensitivity decreased. However, this goes along with a lower level of precision and cannot be generalized to all the datasets, therefore indicating that a lack of power alone is unlikely to explain the detection of less voxels/clusters of MC.

A second factor that potentially influences statistical power is the heterogeneity between experiments (i.e., variations in tasks, stimuli, or control conditions) but also populations (variations in gender and age distribution, recruitment for clinical studies). For effect-size meta-analysis, the power of an analysis is affected by the degree of heterogeneity across studies (Jiang et al., 2010; Kenny & Judd, 2019; McShane & Böckenholt, 2014).

546   Neuroimaging meta-analyses might be similarly affected and with increasing

547   heterogeneity, disproportionately more experiments would be needed to find significant

548   differences in the MC.

549   **Differences in activation strength are lost in MC.** It is important to note that in CM,

550   the factor of interest is modelled as a within-subject factor on the single-experiment

551   level, while for MC the difference between conditions is based on the already sparser

552   representation of the two to-be-compared main effects as peak coordinates. MC

553   analysis thus adds an additional layer on top of classic ALE analyses, resulting in

554   further information reduction. Consequently, the results of meta-analytic contrasts only

555   indirectly include differences in brain activation and no information about effect sizes

556   (Müller, Cieslik, et al., 2018). This sparser representation might affect regions that are

557   involved in both conditions. Indeed, our results suggest that MC is especially insensitive

558   for clusters that are found in both individual meta-analyses that are compared to each

559   other. For example, the MC of emotional face processing where the individual meta-

560   analyses overlap most (33 % of the voxels of the individual meta-analysis emotion >

561   baseline overlap with the voxels of neutral > baseline) reveals only few regions that are

562   also found in CM. In turn, the MC with the highest amount of voxels that are also found

563   in CM (interference processing) exhibited the lowest overlap of 13 % (see

564   supplementary Figure S2-5 for the amount of overlap between the individual meta-

565   analyses of MCs for all datasets). Thus, activation differences for regions involved in

566   both conditions cannot be revealed by MC. However, stronger convergence in similar

567   regions can be observed if regions are larger in extent in one of the individual analyses

568   or are slightly shifted. Therefore, on the regional level, the problem of not detecting

569   differences in convergence in overlapping regions can be mitigated by higher power of

570   the individual meta-analyses, as the clusters become larger, the more experiments are

571   included that show activation differences in a given area (Frahm et al., 2022). This is

572   particularly evident in the WM 2-back > 1-back comparison. The extent of the clusters

573   for 2-back > 0-back found also for 1-back > 0-back is much larger, and MCs can

574   therefore, on the regional level, reveal differences that are also apparent in CMs. The

575 extent and location of convergence thus directly influence if a difference in convergence
576 of a particular contrast can be detected.

577 Less differences in convergence compared to convergence in differences can
578 additionally be attributed to effects driven by deactivations. For example, a region might
579 be slightly activated in condition A and deactivated by condition B. Testing A against
580 baseline as well as B against baseline in an individual fMRI experiment can thus lead to
581 non-significant effects, but testing A against B would reveal significant effects. Thus,
582 regions may be identified in CM but not in the meta-contrast. The converse also holds
583 true, regions observed in individual experiments of A > baseline might be missing in A >
584 B, not due to activations in the control condition B (cancelling out of effects) but due to
585 slight activations in A and deactivations in the baseline condition, leading to significant
586 activation differences in A > baseline. It should be noted that these theoretical
587 considerations also apply to differences in activation in contrasts of single fMRI studies.
588 Previous studies have shown that the activation strength and significance of regions can
589 be directly influenced by the choice of initial conditions at the experiment level (Marx et
590 al., 2004; Newman et al., 2001; Stark & Squire, 2001). This therefore has a direct
591 impact on CM and MC results since activation strength is not considered in the MC
592 calculation and sparsely represented regions will lead to non-significant results on a
593 meta-analysis level. Thus, given that differences in activation is ignored in MC, both
594 deactivations and regions activating on both conditions (overlapping regions) can't be
595 picked up by meta-analytic contrasts.

5.3 Regions showing differences in convergence but no convergence of differences

598 Cluster-wise precision and median peak distance are quite good for the working
599 memory and interference processing datasets, but for emotional face processing there
600 are a lot of regions found in MC but not CM. Thus, for the two datasets from the
601 cognitive domain, there are only a few additional regions in MC, while for the emotional
602 domain the results look different. This can be attributed to systematic experimental and

603 methodological differences between experiments included in the two meta-analyses of
604 MC which are not related to the condition of interest as well as sub optimal cognitive
605 subtraction and thresholding of the original studies. For example, task and stimulus
606 effects that were present in the "emotional > baseline" meta-analysis might have been
607 absent or weaker in the "neutral > baseline" meta-analysis, resulting in task-specific
608 effects not being subtracted in MC. Similarly, general task demands that are unrelated
609 to the process of interest might only become apparent when contrasting against
610 baseline conditions. Thus, divergent effects between MC and CM might be the results of
611 suboptimal cognitive subtraction, especially for baseline contrasts. This becomes
612 especially apparent for the emotion dataset, where primarily regions involved in
613 increased task demands are identified in MC, rather than regions involved in emotional
614 face processing.

615 The emotional face processing dataset shows a lot of variation of the face and control
616 stimuli, as well as in the task performed. Experiments testing the contrast between
617 emotional and neutral faces (which are included in CM) typically use only face stimuli
618 and do not vary the task (e.g., (Seara-Cardoso et al., 2016; Williams et al., 2001)). In
619 contrast, in experiments testing emotional or neutral against control/baseline (which are
620 included in MC), both face and non-face stimuli, and the task is varied for face and
621 control stimuli (i.e., emotion matching for faces and shape matching for control stimuli,
622 e.g., (Binelli et al., 2016; Via et al., 2014)). Together with very similar regions in face-
623 processing-related regions, this could have led to a meta-contrast of more emotion-
624 unspecific and more task-related effects (see supplementary Figure S5). For example,
625 parietal regions, which are found in MC for emotion processing, play a major role in
626 attentional shifts (Langner et al., 2011; Serences et al., 2004; Shomstein & Yantis,
627 2004; Townsend et al., 2006). Therefore, it is quite likely that stronger convergence for
628 the meta-analytic contrast between emotional versus neutral faces reflects those
629 attentional shifts between facial and control stimuli and not necessarily emotional
630 processes. This is supported by the results of the matched analyses where parietal
631 clusters are no longer found for MC when the control condition is matched.

632 In summary, systematic differences and suboptimal cognitive subtraction in experiments

633 used for MC computation may introduce confounds. These are not unique to task-

634 focused meta-analyses but may be even more prevalent in MCs between groups (such

635 as young versus old, patients versus control, disease A versus disease B), since there

636 may be confounding factors in the compared groups that are difficult to control, in

637 addition to the systematic variations in task, control condition and stimuli.

638 Therefore, our results highlight the susceptibility of MC for effects of systematic

639 confounds between the contrasted datasets as well as imperfect cognitive subtraction.

640 These should be considered in the planning, analyses, interpretation and reporting of

641 MC results by applying appropriate inclusion criteria, providing a detailed

642 characterization of the experiments of the datasets, evaluation, and accounting of

643 variables of no interest that may systematically differ between the two meta-analyses as

644 well as transparent reporting.

5.4 Limitations and outlook

646 Some of the analyses presented here were only made possible thanks to additional

647 non-peer-reviewed results sent to us by authors. While this approach introduces a

648 potential new source of variance, it is not uncommon in the meta-analytic literature

649 (Müller, Cieslik, et al., 2018). Furthermore, we were only able to gather all three

650 contrasts of interest (experiments) for a limited number of studies, so we could not

651 eliminate the effects generated by the heterogeneity of experiments. This may have

652 been particularly evident in the emotional face processing comparison. Finally, we

653 restricted ourselves to task fMRI contrasts between different cognitive conditions. Group

654 contrasts, such as patients vs. controls (Costa et al., 2021; Degasperi et al., 2021),

655 were not investigated here. Therefore, caution should be exercised when generalizing

656 the results discussed here to such scenarios, which contain even more sources of

657 heterogeneity.

658 It is important to acknowledge that the individual meta-analyses between which MC was

659 calculated differed in terms of sample size, with condition A (versus baseline) always

660   included more experiments than condition B (versus baseline). This might have

661   influenced the results, given a difference in power between the two analyses. One

662   potential solution would be the use of balanced contrasts (Tamon et al., 2024).

663   However, for this evaluation, we chose to calculate MC in the traditional way, as the

664   results are better generalizable to most applications in the literature.

665   Importantly, although we compared MC to the results of CM and calculated metrics

666   such as sensitivity, we do not claim that CM is the gold standard and free of any bias or

667   problems. That is, systematic confounds that are common across experiments of CM

668   can also lead to convergence in CM that is not necessarily related to the process of

669   interest. Furthermore, as discussed above, coordinate-based meta-analyses use a

670   sparse representation of results of individual studies, with coordinates being strongly

671   dependent on the methodological and reporting characteristics of the individual studies

672   (thresholding, number of peaks reported, etc.). Thus, to allow generalizations and

673   conclusions that go beyond the effects found for CM, MC should ideally be compared to

674   results of the contrast of interest derived from different approaches. This could for

675   example be results based on a large-sample single study (as already done for 2-back >

676   0-back here; reported in the supplementary material) or of an image-based meta-

677   analysis (Salimi-Khorshidi et al., 2009). Unfortunately, the latter is often hampered by a

678   lack of availability of unthresholded maps of individual studies.

679   Lastly, beyond the contrast-analyses used and implemented for ALE meta-analyses,

680   differences between conditions can also be tested by modelling categorical regressors

681   in meta-regression analyses. While this is not currently implemented for ALE, Bayesian

682   spatial regression models (Samartsidis et al., 2019) and coordinate-based meta-

683   regression models (Yu et al., 2024) would allow the modelling of covariates, reflecting a

684   different approach to group comparisons. Although Samartsidis et al. (2019) reported

685   similar results from meta-regression and the contrast analyses of ALE, a comparison of

686   MC to the results of meta-regression would be a valuable extension of the current

687   project.

5.5 Conclusion

ALE meta-analytic contrasts are a widely used method for comparing results between individual meta-analyses and are a necessary extension for meta-analytic studies that claim differences in regional convergence between two analyses across experiments of different conditions, states, or groups. Our results show that these contrasts in most cases capture similar results to standard meta-analyses, but results can suffer from low power, systematic confounds and a sparser representation of effects. Thus, meta-analytic contrasts are beyond demonstrating significant differences in convergence between meta-analytic results, well suited for complementary and exploratory meta-analytic investigations. However, they should not be understood as a direct substitute for classic meta-analyses across experiment-level contrasts reported in the literature, and the regions found should be interpreted accordingly.

6   Data and Code Availability

All ALE meta-analyses were computed using pyALE, accessible at https://github.com/LenFrahm/pyALE. Code for comparing the meta-analytic results and plotting figures can be found at https://github.com/vinkue/ale-meta-contrasts. The data are available upon request, as the coordinates include data from other research groups that have not been consented for redistribution.

The code used to compute the large sample single study contrast was based on the HCPpipelines (https://github.com/Washington-University/HCPpipelines) and the Python exemplary analysis by Esteban et al. (2020) (https://github.com/poldracklab/ds003-post-fMRIPrep-analysis). Data from the Human Connectome Project Young Adults are available at https://db.humanconnectome.org/.

7 Author Contributions

Conceptualization: EC, RL, VM, VK; Software: LF, SE, VK; Formal analysis: LF, VK; Data Curation: VK, VM; Writing - Original Draft: VM, VK; Writing - Review & Editing: EC, FH, LF, RL, SE, VM, VK; Visualization: VK; Supervision: RL, SE, VM.

8 Declaration of Competing Interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

9 Acknowledgements

References

740 Adolphs, R. (2002). Neural systems for recognizing emotion. *Current Opinion in*

741 *Neurobiology*, *12*(2), 169–177. https://doi.org/10.1016/S0959-4388(02)00301-X

742 Agostini, A., Ballotta, D., Righi, S., Moretti, M., Bertani, A., Scarcelli, A., Sartini, A.,

743 Ercolani, M., Nichelli, P., Campieri, M., & Benuzzi, F. (2017). Stress and brain functional

744 changes in patients with Crohn's disease: A functional magnetic resonance imaging

745 study. *Neurogastroenterology & Motility*, *29*(10), e13108.

746 https://doi.org/10.1111/nmo.13108

747 Aguirre, N., Cruz-Gómez, Á. J., Miró-Padilla, A., Bueichekú, E., Broseta Torres, R.,

748 Ávila, C., Sanchis-Segura, C., & Forn, C. (2019). Repeated Working Memory Training

749 Improves Task Performance and Neural Efficiency in Multiple Sclerosis Patients and

750 Healthy Controls. *Multiple Sclerosis International*, *2019*, 1–13.

751 https://doi.org/10.1155/2019/2657902

752 Binelli, C., Muñiz, A., Subira, S., Navines, R., Blanco-Hinojo, L., Perez-Garcia, D.,

753 Crippa, J., Farré, M., Pérez-Jurado, L., Pujol, J., & Martin-Santos, R. (2016). Facial

754 emotion processing in patients with social anxiety disorder and Williams–Beuren

755 syndrome: An fMRI study. *Journal of Psychiatry and Neuroscience*, *41*(3), 182–191.

756 https://doi.org/10.1503/jpn.140384

757 Bossier, H., Roels, S. P., Seurinck, R., Banaschewski, T., Barker, G. J., Bokde, A. L.

758 W., Quinlan, E. B., Desrivières, S., Flor, H., Grigis, A., Garavan, H., Gowland, P., Heinz,

759 A., Ittermann, B., Martinot, J.-L., Artiges, E., Nees, F., Orfanos, D. P., Poustka, L., …

760 Moerkerke, B. (2020). The empirical replicability of task-based fMRI as a function of

761 sample size. *NeuroImage*, *212*, 116601.

762 https://doi.org/10.1016/j.neuroimage.2020.116601

763 Camilleri, J. A., Müller, V. I., Fox, P., Laird, A. R., Hoffstaedter, F., Kalenscher, T., &

764 Eickhoff, S. B. (2018). Definition and characterization of an extended multiple-demand

765 network. *NeuroImage*, *165*, 138–147. https://doi.org/10.1016/j.neuroimage.2017.10.020

766 Campanella, S., Peigneux, P., Petit, G., Lallemand, F., Saeremans, M., Noël, X.,

767 Metens, T., Nouali, M., De Tiège, X., De Witte, P., Ward, R., & Verbanck, P. (2013).

768 Increased Cortical Activity in Binge Drinkers during Working Memory Task: A

769 Preliminary Assessment through a Functional Magnetic Resonance Imaging Study.

770 *PLoS ONE*, *8*(4), e62260. https://doi.org/10.1371/journal.pone.0062260

771 Caspers, S., Zilles, K., Laird, A. R., & Eickhoff, S. B. (2010). ALE meta-analysis of

772 action observation and imitation in the human brain. *NeuroImage*, *50*(3), 1148–1167.

773 https://doi.org/10.1016/j.neuroimage.2009.12.112

774 Cieslik, E. C., Mueller, V. I., Eickhoff, C. R., Langner, R., & Eickhoff, S. B. (2015). Three

775 key regions for supervisory attentional control: Evidence from neuroimaging meta-

776 analyses. *Neuroscience & Biobehavioral Reviews*, *48*, 22–34.

777 https://doi.org/10.1016/j.neubiorev.2014.11.003

778 Cole, M. W., & Schneider, W. (2007). The cognitive control network: Integrated cortical

779 regions with dissociable functions. *NeuroImage*, *37*(1), 343–360.

780 https://doi.org/10.1016/j.neuroimage.2007.03.071

781 Costa, C., Cristea, I. A., Dal Bò, E., Melloni, C., & Gentili, C. (2021). Brain activity during

782 facial processing in autism spectrum disorder: An activation likelihood estimation (ALE)

783 meta-analysis of neuroimaging studies. *Journal of Child Psychology and Psychiatry*,

784 *62*(12), 1412–1424. https://doi.org/10.1111/jcpp.13412

785 Cullen, K. R., LaRiviere, L. L., Vizueta, N., Thomas, K. M., Hunt, R. H., Miller, M. J.,

786 Lim, K. O., & Schulz, S. C. (2016). Brain activation in response to overt and covert fear

787 and happy faces in women with borderline personality disorder. *Brain Imaging and*

788 *Behavior*, *10*(2), 319–331. https://doi.org/10.1007/s11682-015-9406-4

789   Daamen, M., Bäuml, J. G., Scheef, L., Sorg, C., Busch, B., Baumann, N., Bartmann, P.,

790   Wolke, D., Wohlschläger, A., & Boecker, H. (2015). Working memory in preterm-born

791   adults: Load-dependent compensatory activity of the posterior default mode network:

792   Working Memory in Preterm-Born Adults. *Human Brain Mapping*, *36*(3), 1121–1137.

793   https://doi.org/10.1002/hbm.22691

794   Dan, R., Růžička, F., Bezdicek, O., Roth, J., Růžička, E., Vymazal, J., Goelman, G., &

795   Jech, R. (2019). Impact of dopamine and cognitive impairment on neural reactivity to

796   facial emotion in Parkinson's disease. *European Neuropsychopharmacology*, *29*(11),

797   1258–1272. https://doi.org/10.1016/j.euroneuro.2019.09.003

798   Degasperi, G., Cristea, I. A., Di Rosa, E., Costa, C., & Gentili, C. (2021). Parsing

799   variability in borderline personality disorder: A meta-analysis of neuroimaging studies.

800   *Translational Psychiatry*, *11*(1), 314. https://doi.org/10.1038/s41398-021-01446-z

801   Dolcos, F., Iordan, A. D., & Dolcos, S. (2011). Neural correlates of emotion-cognition

802   interactions: A review of evidence from brain imaging investigations. *Journal of*

803   *Cognitive Psychology (Hove, England)*, *23*(6), 669–694.

804   https://doi.org/10.1080/20445911.2011.594433

805   Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: Mental

806   programs for intelligent behaviour. *Trends in Cognitive Sciences*, *14*(4), 172–179.

807   https://doi.org/10.1016/j.tics.2010.01.004

808   Eickhoff, S. B., Bzdok, D., Laird, A. R., Kurth, F., & Fox, P. T. (2012). Activation

809   likelihood estimation meta-analysis revisited. *NeuroImage*, *59*(3), 2349–2361.

810   https://doi.org/10.1016/j.neuroimage.2011.09.017

811   Eickhoff, S. B., Bzdok, D., Laird, A. R., Roski, C., Caspers, S., Zilles, K., & Fox, P. T.

812   (2011). Co-activation patterns distinguish cortical modules, their connectivity and

813   functional differentiation. *NeuroImage*, *57*(3), 938–949.

814   https://doi.org/10.1016/j.neuroimage.2011.05.021

815 Eickhoff, S. B., Laird, A. R., Grefkes, C., Wang, L. E., Zilles, K., & Fox, P. T. (2009).

816 Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data:

817 A random-effects approach based on empirical estimates of spatial uncertainty. *Human*

818 *Brain Mapping*, *30*(9), 2907–2926. https://doi.org/10.1002/hbm.20718

819 Eickhoff, S. B., Nichols, T. E., Laird, A. R., Hoffstaedter, F., Amunts, K., Fox, P. T.,

820 Bzdok, D., & Eickhoff, C. R. (2016). Behavior, sensitivity, and power of activation

821 likelihood estimation characterized by massive empirical simulation. *NeuroImage*, *137*,

822 70–85. https://doi.org/10.1016/j.neuroimage.2016.04.072

823 Esteban, O., Ciric, R., Finc, K., Blair, R. W., Markiewicz, C. J., Moodie, C. A., Kent, J.

824 D., Goncalves, M., DuPre, E., Gomez, D. E. P., Ye, Z., Salo, T., Valabregue, R., Amlien,

825 I. K., Liem, F., Jacoby, N., Stojić, H., Cieslak, M., Urchs, S., … Gorgolewski, K. J.

826 (2020). Analysis of task-based functional MRI data preprocessed with fMRIPrep. *Nature*

827 *Protocols*, *15*(7), 2186–2202. https://doi.org/10.1038/s41596-020-0327-3

828 Fehlbaum, L. V., Borbás, R., Paul, K., Eickhoff, S. b, & Raschle, N. m. (2021). Early and

829 late neural correlates of mentalizing: ALE meta-analyses in adults, children and

830 adolescents. *Social Cognitive and Affective Neuroscience*, nsab105.

831 https://doi.org/10.1093/scan/nsab105

832 Frahm, L., Cieslik, E. C., Hoffstaedter, F., Satterthwaite, T. D., Fox, P. T., Langner, R.,

833 & Eickhoff, S. B. (2022). Evaluation of thresholding methods for activation likelihood

834 estimation meta-analysis via large-scale simulations. *Human Brain Mapping*, *43*(13),

835 3987–3997. https://doi.org/10.1002/hbm.25898

836 Fukuda, Y., Katthagen, T., Deserno, L., Shayegan, L., Kaminski, J., Heinz, A., &

837 Schlagenhauf, F. (2019). Reduced parietofrontal effective connectivity during a working-

838 memory task in people with high delusional ideation. *Journal of Psychiatry &*

839 *Neuroscience*, *44*(3), 195–204. https://doi.org/10.1503/jpn.180043

840 Gan, X., Zhou, X., Li, J., Jiao, G., Jiang, X., Biswal, B., Yao, S., Klugah-Brown, B., &

841 Becker, B. (2022). Common and distinct neurofunctional representations of core and

842 social disgust in the brain: Coordinate-based and network meta-analyses. *Neuroscience*

843 *& Biobehavioral Reviews*, *135*, 104553. https://doi.org/10.1016/j.neubiorev.2022.104553

844 Ghavidel, N., Khodagholi, F., Ahmadiani, A., Khosrowabadi, R., Asadi, S., & Shams, J.

845 (2020). Frontocingulate Dysfunction Is Associated with Depression and Decreased

846 Serum PON1 in Methamphetamine-Dependent Patients. *Neuropsychiatric Disease and*

847 *Treatment*, *Volume 16*, 489–499. https://doi.org/10.2147/NDT.S237528

848 Habel, U., Koch, K., Pauly, K., Kellermann, T., Reske, M., Backes, V., Seiferth, N. Y.,

849 Stöcker, T., Kircher, T., Amunts, K., Jon Shah, N., & Schneider, F. (2007). The influence

850 of olfactory-induced negative emotion on verbal working memory: Individual differences

851 in neurobehavioral findings. *Brain Research*, *1152*, 158–170.

852 https://doi.org/10.1016/j.brainres.2007.03.048

853 Harding, I. H., Corben, L. A., Storey, E., Egan, G. F., Stagnitti, M. R., Poudel, G. R.,

854 Delatycki, M. B., & Georgiou-Karistianis, N. (2016). Fronto-cerebellar dysfunction and

855 dysconnectivity underlying cognition in friedreich ataxia: The IMAGE-FRDA study:

856 Cognitive Networks in Friedreich Ataxia. *Human Brain Mapping*, *37*(1), 338–350.

857 https://doi.org/10.1002/hbm.23034

858 Hill, A. C., Laird, A. R., & Robinson, J. L. (2014). Gender differences in working memory

859 networks: A BrainMap meta-analysis. *Biological Psychology*, *102*, 18–29.

860 https://doi.org/10.1016/j.biopsycho.2014.06.008

861 Jaccard, P. (1901). *Étude comparative de la distribution florale dans une portion des*

862 *Alpes et du Jura*. https://doi.org/10.5169/SEALS-266450

863 Jiang, D., Pepler, D., & Yao, H. (2010). The effect of population heterogeneity on

864 statistical power in the design and evaluation of interventions. *International Journal of*

865 *Behavioral Development*, *34*(5), 473–480. https://doi.org/10.1177/0165025410375912

866 Jung, K., Friston, K. J., Pae, C., Choi, H. H., Tak, S., Choi, Y. K., Park, B., Park, C.-A.,

867 Cheong, C., & Park, H.-J. (2018). Effective connectivity during working memory and

868   resting states: A DCM study. *NeuroImage, 169*, 485–495.

869   https://doi.org/10.1016/j.neuroimage.2017.12.067

870   Kaminski, J., Gleich, T., Fukuda, Y., Katthagen, T., Gallinat, J., Heinz, A., &

871   Schlagenhauf, F. (2020). Association of Cortical Glutamate and Working Memory

872   Activation in Patients With Schizophrenia: A Multimodal Proton Magnetic Resonance

873   Spectroscopy and Functional Magnetic Resonance Imaging Study. *Biological*

874   *Psychiatry*, *87*(3), 225–233. https://doi.org/10.1016/j.biopsych.2019.07.011

875   Kenny, D. A., & Judd, C. M. (2019). The unappreciated heterogeneity of effect sizes:

876   Implications for power, precision, planning of research, and replication. *Psychological*

877   *Methods*, *24*(5), 578–589. https://doi.org/10.1037/met0000209

878   King, T. Z., Na, S., & Mao, H. (2015). Neural Underpinnings of Working Memory in Adult

879   Survivors of Childhood Brain Tumors. *Journal of the International Neuropsychological*

880   *Society*, *21*(7), 494–505. https://doi.org/10.1017/S135561771500051X

881   Klugah-Brown, B., Zhou, X., Pradhan, B. K., Zweerings, J., Mathiak, K., Biswal, B., &

882   Becker, B. (2021). Common neurofunctional dysregulations characterize obsessive–

883   compulsive, substance use, and gaming disorders—An activation likelihood meta-

884   analysis of functional imaging studies. *Addiction Biology*, *26*(4).

885   https://doi.org/10.1111/adb.12997

886   Kogler, L., Müller, V. I., Chang, A., Eickhoff, S. B., Fox, P. T., Gur, R. C., & Derntl, B.

887   (2015). Psychosocial versus physiological stress - Meta-analyses on deactivations and

888   activations of the neural correlates of stress reactions. *NeuroImage, 119,* 235–251.

889   https://doi.org/10.1016/j.neuroimage.2015.06.059

890   Kogler, L., Müller, V. I., Werminghausen, E., Eickhoff, S. B., & Derntl, B. (2020). Do I

891   feel or do I know? Neuroimaging meta-analyses on the multiple facets of empathy.

892   *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *129*, 341–

893   355. https://doi.org/10.1016/j.cortex.2020.04.031

894 Köhler, S., Bär, K.-J., & Wagner, G. (2016). Differential involvement of brainstem

895 noradrenergic and midbrain dopaminergic nuclei in cognitive control: NA and

896 Dopaminergic Nuclei and Cognitive Control. *Human Brain Mapping*, *37*(6), 2305–2318.

897 https://doi.org/10.1002/hbm.23173

898 Kowalczyk, O. S., Pauls, A. M., Fusté, M., Williams, S. C. R., Hazelgrove, K., Vecchio,

899 C., Seneviratne, G., Pariante, C. M., Dazzan, P., & Mehta, M. A. (2021). Neurocognitive

900 correlates of working memory and emotional processing in postpartum psychosis: An

901 fMRI study. *Psychological Medicine*, *51*(10), 1724–1732.

902 https://doi.org/10.1017/S0033291720000471

903 Kozasa, E. H., Balardin, J. B., Sato, J. R., Chaim, K. T., Lacerda, S. S., Radvany, J.,

904 Mello, L. E. A. M., & Amaro, E. (2018). Effects of a 7-Day Meditation Retreat on the

905 Brain Function of Meditators and Non-Meditators During an Attention Task. *Frontiers in*

906 *Human Neuroscience*, *12*, 222. https://doi.org/10.3389/fnhum.2018.00222

907 Kronbichler, L., Stelzig-Schöler, R., Pearce, B.-G., Tschernegg, M., Said-Yürekli, S.,

908 Reich, L. A., Weber, S., Aichhorn, W., & Kronbichler, M. (2018). Schizophrenia and

909 Category-Selectivity in the Brain: Normal for Faces but Abnormal for Houses. *Frontiers*

910 *in Psychiatry*, *9*, 47. https://doi.org/10.3389/fpsyt.2018.00047

911 Krug, A., Markov, V., Eggermann, T., Krach, S., Zerres, K., Stöcker, T., Shah, N. J.,

912 Schneider, F., Nöthen, M. M., Treutlein, J., Rietschel, M., & Kircher, T. (2008). Genetic

913 variation in the schizophrenia-risk gene Neuregulin1 correlates with differences in

914 frontal brain activation in a working memory task in healthy individuals. *NeuroImage*,

915 *42*(4), 1569–1576. https://doi.org/10.1016/j.neuroimage.2008.05.058

916 Lahr, J., Minkova, L., Tabrizi, S. J., Stout, J. C., Klöppel, S., Scheller, E., & the

917 TrackOn-HD Investigators. (2018). Working Memory-Related Effective Connectivity in

918 Huntington's Disease Patients. *Frontiers in Neurology*, *9*, 370.

919 https://doi.org/10.3389/fneur.2018.00370

920 Laird, A. R., Fox, P. M., Price, C. J., Glahn, D. C., Uecker, A. M., Lancaster, J. L.,

921 Turkeltaub, P. E., Kochunov, P., & Fox, P. T. (2005). ALE meta-analysis: Controlling the

922 false discovery rate and performing statistical contrasts. *Human Brain Mapping*, *25*(1),

923 155–164. https://doi.org/10.1002/hbm.20136

924 Langner, R., & Eickhoff, S. B. (2013). Sustaining attention to simple tasks: A meta-

925 analytic review of the neural mechanisms of vigilant attention. *Psychological Bulletin*,

926 *139*(4), 870–900. https://doi.org/10.1037/a0030694

927 Langner, R., Kellermann, T., Boers, F., Sturm, W., Willmes, K., & Eickhoff, S. B. (2011).

928 Modality-Specific Perceptual Expectations Selectively Modulate Baseline Activity in

929 Auditory, Somatosensory, and Visual Cortices. *Cerebral Cortex*, *21*(12), 2850–2862.

930 https://doi.org/10.1093/cercor/bhr083

931 Langner, R., Leiberg, S., Hoffstaedter, F., & Eickhoff, S. B. (2018). Towards a human

932 self-regulation system: Common and distinct neural signatures of emotional and

933 behavioural control. *Neuroscience & Biobehavioral Reviews*, *90*, 400–410.

934 https://doi.org/10.1016/j.neubiorev.2018.04.022

935 Li, L., Zhang, S., Cui, J., Chen, L.-Z., Wang, X., Fan, M., & Wei, G.-X. (2019). Fitness-

936 Dependent Effect of Acute Aerobic Exercise on Executive Function. *Frontiers in*

937 *Physiology*, *10*, 902. https://doi.org/10.3389/fphys.2019.00902

938 Luethi, M. S., Friese, M., Binder, J., Boesiger, P., Luechinger, R., & Rasch, B. (2016).

939 Motivational incentives lead to a strong increase in lateral prefrontal activity after self-

940 control exertion. *Social Cognitive and Affective Neuroscience*, *11*(10), 1618–1626.

941 https://doi.org/10.1093/scan/nsw073

942 Maitra, R. (2010). A re-defined and generalized percent-overlap-of-activation measure

943 for studies of fMRI reproducibility and its use in identifying outlier activation maps.

944 *NeuroImage*, *50*(1), 124–135. https://doi.org/10.1016/j.neuroimage.2009.11.070

945 Marx, E., Deutschländer, A., Stephan, T., Dieterich, M., Wiesmann, M., & Brandt, T.

946 (2004). Eyes open and eyes closed as rest conditions: Impact on brain activation

947 patterns. *NeuroImage*, *21*(4), 1818–1824.

948 https://doi.org/10.1016/j.neuroimage.2003.12.026

949 McShane, B. B., & Böckenholt, U. (2014). You Cannot Step Into the Same River Twice:

950 When Power Analyses Are Optimistic. *Perspectives on Psychological Science*, *9*(6),

951 612–625. https://doi.org/10.1177/1745691614548513

952 Menon, V. (2011). Large-scale brain networks and psychopathology: A unifying triple

953 network model. *Trends in Cognitive Sciences*, *15*(10), 483–506.

954 https://doi.org/10.1016/j.tics.2011.08.003

955 Miró-Padilla, A., Bueichekú, E., Ventura-Campos, N., Flores-Compañ, M.-J., Parcet, M.

956 A., & Ávila, C. (2019). Long-term brain effects of N-back training: An fMRI study. *Brain*

957 *Imaging and Behavior*, *13*(4), 1115–1127. https://doi.org/10.1007/s11682-018-9925-x

958 Müller, V. I., Cieslik, E. C., Laird, A. R., Fox, P. T., Radua, J., Mataix-Cols, D., Tench, C.

959 R., Yarkoni, T., Nichols, T. E., Turkeltaub, P. E., Wager, T. D., & Eickhoff, S. B. (2018).

960 Ten simple rules for neuroimaging meta-analysis. *Neuroscience & Biobehavioral*

961 *Reviews*, *84*, 151–161. https://doi.org/10.1016/j.neubiorev.2017.11.012

962 Müller, V. I., Cieslik, E. C., Serbanescu, I., Laird, A. R., Fox, P. T., & Eickhoff, S. B.

963 (2017). Altered Brain Activity in Unipolar Depression Revisited: Meta-analyses of

964 Neuroimaging Studies. *JAMA Psychiatry*, *74*(1), 47.

965 https://doi.org/10.1001/jamapsychiatry.2016.2783

966 Müller, V. I., Höhner, Y., & Eickhoff, S. B. (2018). Influence of task instructions and

967 stimuli on the neural network of face processing: An ALE meta-analysis. *Cortex*, *103*,

968 240–255. https://doi.org/10.1016/j.cortex.2018.03.011

969 Newman, S. D., Twieg, D. B., & Carpenter, P. A. (2001). Baseline conditions and

970 subtractive logic in neuroimaging. *Human Brain Mapping*, *14*(4), 228–235.

971 https://doi.org/10.1002/hbm.1055

972 Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working

973 memory paradigm: A meta-analysis of normative functional neuroimaging studies.

974 *Human Brain Mapping*, *25*(1), 46–59. https://doi.org/10.1002/hbm.20131

975 Papalini, S., Michels, F., Kohn, N., Wegman, J., van Hemert, S., Roelofs, K., Arias-

976 Vasquez, A., & Aarts, E. (2019). Stress matters: Randomized controlled trial on the

977 effect of probiotics on neurocognition. *Neurobiology of Stress*, *10*, 100141.

978 https://doi.org/10.1016/j.ynstr.2018.100141

979 Peven, J. C., Litz, G. A., Brown, B., Xie, X., Grove, G. A., Watt, J. C., & Erickson, K. I.

980 (2019). Higher Cardiorespiratory Fitness is Associated with Reduced Functional Brain

981 Connectivity During Performance of the Stroop Task. *Brain Plasticity*, *5*(1), 57–67.

982 https://doi.org/10.3233/BPL-190085

983 Phillips, M. L., Drevets, W. C., Rauch, S. L., & Lane, R. (2003). Neurobiology of emotion

984 perception I: The neural basis of normal emotion perception. *Biological Psychiatry*,

985 *54*(5), 504–514. https://doi.org/10.1016/s0006-3223(03)00168-9

986 Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A. R., Schulz, J. B., Fox, P. T., &

987 Eickhoff, S. B. (2012). Modelling neural correlates of working memory: A coordinate-

988 based meta-analysis. *NeuroImage*, *60*(1), 830–846.

989 https://doi.org/10.1016/j.neuroimage.2011.11.050

990 Salimi-Khorshidi, G., Smith, S. M., Keltner, J. R., Wager, T. D., & Nichols, T. E. (2009).

991 Meta-analysis of neuroimaging data: A comparison of image-based and coordinate-

992 based pooling of studies. *NeuroImage*, *45*(3), 810–823.

993 https://doi.org/10.1016/j.neuroimage.2008.12.039

994 Samartsidis, P., Eickhoff, C. R., Eickhoff, S. B., Wager, T. D., Barrett, L. F., Atzil, S.,

995 Johnson, T. D., & Nichols, T. E. (2019). Bayesian Log-Gaussian Cox Process

996 Regression: Applications to Meta-Analysis of Neuroimaging Working Memory Studies.

997 *Journal of the Royal Statistical Society Series C: Applied Statistics*, *68*(1), 217–234.

998 https://doi.org/10.1111/rssc.12295

999   Schlagenhauf, F., Wüstenberg, T., Schmack, K., Dinges, M., Wrase, J., Koslowski, M.,

1000   Kienast, T., Bauer, M., Gallinat, J., Juckel, G., & Heinz, A. (2008). Switching

1001   schizophrenia patients from typical neuroleptics to olanzapine: Effects on BOLD

1002   response during attention and working memory. *European Neuropsychopharmacology*,

1003   *18*(8), 589–599. https://doi.org/10.1016/j.euroneuro.2008.04.013

1004   Schmidt, C., Peigneux, P., Leclercq, Y., Sterpenich, V., Vandewalle, G., Phillips, C.,

1005   Berthomier, P., Berthomier, C., Tinguely, G., Gais, S., Schabus, M., Desseilles, M.,

1006   Dang-Vu, T., Salmon, E., Degueldre, C., Balteau, E., Luxen, A., Cajochen, C., Maquet,

1007   P., & Collette, F. (2012). Circadian Preference Modulates the Neural Substrate of

1008   Conflict Processing across the Day. *PLoS ONE*, *7*(1), e29658.

1009   https://doi.org/10.1371/journal.pone.0029658

1010   Schneider, F., Habel, U., Reske, M., Kellermann, T., Stöcker, T., Shah, N. J., Zilles, K.,

1011   Braus, D. F., Schmitt, A., Schlösser, R., Wagner, M., Frommann, I., Kircher, T., Rapp,

1012   A., Meisenzahl, E., Ufer, S., Ruhrmann, S., Thienel, R., Sauer, H., … Gaebel, W.

1013   (2007). Neural correlates of working memory dysfunction in first-episode schizophrenia

1014   patients: An fMRI multi-center study. *Schizophrenia Research*, *89*(1-3), 198–210.

1015   https://doi.org/10.1016/j.schres.2006.07.021

1016   Seara-Cardoso, A., Sebastian, C. L., Viding, E., & Roiser, J. P. (2016). Affective

1017   resonance in response to others' emotional faces varies with affective ratings and

1018   psychopathic traits in amygdala and anterior insula. *Social Neuroscience*, *11*(2), 140–

1019   152. https://doi.org/10.1080/17470919.2015.1044672

1020   Serences, J. T., Schwarzbach, J., Courtney, S. M., Golay, X., & Yantis, S. (2004).

1021   Control of Object-based Attention in Human Cortex. *Cerebral Cortex*, *14*(12), 1346–

1022   1357. https://doi.org/10.1093/cercor/bhh095

1023   Shashidhara, S., Spronkers, F. S., & Erez, Y. (2020). Individual-subject Functional

1024   Localization Increases Univariate Activation but Not Multivariate Pattern Discriminability

1025   in the "Multiple-demand" Frontoparietal Network. *Journal of Cognitive Neuroscience*,

1026   *32*(7), 1348–1368. https://doi.org/10.1162/jocn_a_01554

1027   Shomstein, S., & Yantis, S. (2004). Control of Attention Shifts between Vision and

1028   Audition in Human Cortex. *The Journal of Neuroscience*, *24*(47), 10702.

1029   https://doi.org/10.1523/JNEUROSCI.2939-04.2004

1030   Smits, M., Dippel, D. W. J., Houston, G. C., Wielopolski, P. A., Koudstaal, P. J., Hunink,

1031   M. G. M., & van der Lugt, A. (2009). Postconcussion syndrome after minor head injury:

1032   Brain activation of working memory and attention. *Human Brain Mapping*, *30*(9), 2789–

1033   2803. https://doi.org/10.1002/hbm.20709

1034   Stark, C. E. L., & Squire, L. R. (2001). When zero is not zero: The problem of

1035   ambiguous baseline conditions in fMRI. *Proceedings of the National Academy of*

1036   *Sciences*, *98*(22), 12760–12766. https://doi.org/10.1073/pnas.221462998

1037   Stevens, J. S., & Hamann, S. (2012). Sex differences in brain activation to emotional

1038   stimuli: A meta-analysis of neuroimaging studies. *Neuropsychologia*, *50*(7), 1578–1593.

1039   https://doi.org/10.1016/j.neuropsychologia.2012.03.011

1040   Swick, D., Ashley, V., & Turken, U. (2011). Are the neural correlates of stopping and not

1041   going identical? Quantitative meta-analysis of two response inhibition tasks.

1042   *NeuroImage*, *56*(3), 1655–1665. https://doi.org/10.1016/j.neuroimage.2011.02.070

1043   Szabó, E., Galambos, A., Kocsel, N., Édes, A. E., Pap, D., Zsombók, T., Kozák, L. R.,

1044   Bagdy, G., Kökönyei, G., & Juhász, G. (2019). Association between migraine frequency

1045   and neural response to emotional faces: An fMRI study. *NeuroImage: Clinical*, *22*,

1046   101790. https://doi.org/10.1016/j.nicl.2019.101790

1047   Takeuchi, H., Taki, Y., Nouchi, R., Yokoyama, R., Kotozaki, Y., Nakagawa, S.,

1048   Sekiguchi, A., Iizuka, K., Hanawa, S., Araki, T., Miyauchi, C. M., Sakaki, K., Sassa, Y.,

1049   Nozawa, T., Ikeda, S., Yokota, S., Daniele, M., & Kawashima, R. (2018). General

1050   intelligence is associated with working memory-related brain activity: New evidence

1051   from a large sample study. *Brain Structure and Function*, *223*(9), 4243–4258.

1052   https://doi.org/10.1007/s00429-018-1747-5

1053 Tamon, H., Fujino, J., Itahashi, T., Frahm, L., Parlatini, V., Aoki, Y. Y., Castellanos, F.
1054 X., Eickhoff, S. B., & Cortese, S. (2024). Shared and Specific Neural Correlates of
1055 Attention Deficit Hyperactivity Disorder and Autism Spectrum Disorder: A Meta-Analysis
1056 of 243 Task-Based Functional MRI Studies. *American Journal of Psychiatry*, *181*(6),
1057 541–552. https://doi.org/10.1176/appi.ajp.20230270

1058 Townsend, J., Adamo, M., & Haist, F. (2006). Changing channels: An fMRI study of
1059 aging and cross-modal attention shifts. *NeuroImage*, *31*(4), 1682–1692.
1060 https://doi.org/10.1016/j.neuroimage.2006.01.045

1061 Turkeltaub, P. E., Eden, G. F., Jones, K. M., & Zeffiro, T. A. (2002). Meta-analysis of the
1062 functional neuroanatomy of single-word reading: Method and validation. *NeuroImage*,
1063 *16*(3 Pt 1), 765–780. https://doi.org/10.1006/nimg.2002.1131

1064 Turkeltaub, P. E., Eickhoff, S. B., Laird, A. R., Fox, M., Wiener, M., & Fox, P. (2012).
1065 Minimizing within-experiment and within-group effects in activation likelihood estimation
1066 meta-analyses. *Human Brain Mapping*, *33*(1), 1–13. https://doi.org/10.1002/hbm.21186

1067 Vacchi, L., Rocca, M. A., Meani, A., Rodegher, M., Martinelli, V., Comi, G., Falini, A., &
1068 Filippi, M. (2017). Working memory network dysfunction in relapse-onset multiple
1069 sclerosis phenotypes: A clinical-imaging evaluation. *Multiple Sclerosis Journal*, *23*(4),
1070 577–587. https://doi.org/10.1177/1352458516656809

1071 van der Horn, H. J., Liemburg, E. J., Scheenen, M. E., de Koning, M. E., Spikman, J.
1072 M., & van der Naalt, J. (2016). Post-concussive complaints after mild traumatic brain
1073 injury associated with altered brain networks during working memory performance.
1074 *Brain Imaging and Behavior*, *10*(4), 1243–1253. https://doi.org/10.1007/s11682-015-
1075 9489-y

1076 Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., & Ugurbil,
1077 K. (2013). The WU-Minn Human Connectome Project: An overview. *NeuroImage*, *80*,
1078 62–79. https://doi.org/10.1016/j.neuroimage.2013.05.041

1079   Via, E., Cardoner, N., Pujol, J., Alonso, P., López-Solà, M., Real, E., Contreras-

1080   Rodríguez, O., Deus, J., Segalàs, C., Menchón, J. M., Soriano-Mas, C., & Harrison, B.

1081   J. (2014). Amygdala activation and symptom dimensions in obsessive–compulsive

1082   disorder. *British Journal of Psychiatry*, *204*(1), 61–68.

1083   https://doi.org/10.1192/bjp.bp.112.123364

1084   Wagner, G., De la Cruz, F., Schachtzabel, C., Güllmar, D., Schultz, C. C., Schlösser, R.

1085   G., Bär, K.-J., & Koch, K. (2015). Structural and functional dysconnectivity of the fronto-

1086   thalamic system in schizophrenia: A DCM-DTI study. *Cortex*, *66*, 35–45.

1087   https://doi.org/10.1016/j.cortex.2015.02.004

1088   Wallentin, M., Gravholt, C. H., & Skakkebæk, A. (2015). Broca's region and Visual Word

1089   Form Area activation differ during a predictive Stroop task. *Cortex*, *73*, 257–270.

1090   https://doi.org/10.1016/j.cortex.2015.08.023

1091   Wesley, M. J., & Bickel, W. K. (2014). Remember the Future II: Meta-analyses and

1092   Functional Overlap of Working Memory and Delay Discounting. *Biological Psychiatry*,

1093   *75*(6), 435–448. https://doi.org/10.1016/j.biopsych.2013.08.008

1094   Williams, L. M., Phillips, M. L., Brammer, M. J., Skerrett, D., Lagopoulos, J., Rennie, C.,

1095   Bahramali, H., Olivieri, G., David, A. S., Peduto, A., & Gordon, E. (2001). Arousal

1096   Dissociates Amygdala and Hippocampal Fear Responses: Evidence from Simultaneous

1097   fMRI and Skin Conductance Recording. *NeuroImage*, *14*(5), 1070–1079.

1098   https://doi.org/10.1006/nimg.2001.0904

1099   Wu, S., Wang, H., Chen, C., Zou, J., Huang, H., Li, P., Zhao, Y., Xu, Q., Zhang, L.,

1100   Wang, H., Pandit, S., Dahal, S., Chen, J., Zhou, Y., Jiang, T., & Wang, G. (2017). Task

1101   Performance Modulates Functional Connectivity Involving the Dorsolateral Prefrontal

1102   Cortex in Patients with Schizophrenia. *Frontiers in Psychology*, *8*.

1103   https://doi.org/10.3389/fpsyg.2017.00056

1104   Yaple, Z. A., Stevens, W. D., & Arsalidou, M. (2019). Meta-analyses of the n-back

1105   working memory task: fMRI evidence of age-related changes in prefrontal cortex

1106   involvement across the adult lifespan. *NeuroImage*, *196*, 16–31.

1107   https://doi.org/10.1016/j.neuroimage.2019.03.074

1108   Ye, Z., & Zhou, X. (2009). Conflict control during sentence comprehension: fMRI

1109   evidence. *NeuroImage*, *48*(1), 280–290.

1110   https://doi.org/10.1016/j.neuroimage.2009.06.032

1111   Yu, Y., Lobo, R. P., Riedel, M. C., Bottenhorn, K., Laird, A. R., & Nichols, T. E. (2024).

1112   Neuroimaging meta regression for coordinate based meta analysis data with a spatial

1113   model. *Biostatistics*, kxae024. https://doi.org/10.1093/biostatistics/kxae024

1114   Zhang, Z., Peng, P., Eickhoff, S. B., Lin, X., Zhang, D., & Wang, Y. (2021). Neural

1115   substrates of the executive function construct, age-related changes, and task materials

1116   in adolescents and adults: ALE meta-analyses of 408 fMRI studies. *Developmental*

1117   *Science*, desc.13111. https://doi.org/10.1111/desc.13111