

Supplementary Material—Meta-Analysis of the Functional Neuroimaging Literature with Probabilistic Logic Programming

Majd Abdallah^{1,+}, Valentin Iovene^{1,+}, Gaston Zanitti¹, and Demian Wassermann^{1,*}

¹Inria, CEA, Neurospin, MIND team, Université Paris Saclay, 91120 Palaiseau, France

*Corresponding author: demian.wassermann@inria.fr

ABSTRACT

Inferring reliable brain-behavior associations requires synthesizing evidence from thousands of functional neuroimaging studies through meta-analysis. However, existing meta-analysis tools are limited to investigating simple neuroscience concepts and expressing a restricted range of questions. Here, we expand the scope of neuroimaging meta-analysis by designing NeuroLang: a domain-specific language to express and test hypotheses using probabilistic first-order logic programming. By leveraging formalisms found at the crossroads of artificial intelligence and knowledge representation, NeuroLang provides the expressivity to address a larger repertoire of hypotheses in a meta-analysis, while seamlessly modeling the uncertainty inherent to neuroimaging data. We demonstrate the language’s capabilities in conducting comprehensive neuroimaging meta-analysis through use-case examples that address questions of structure-function associations. Specifically, we infer the specific functional roles of three canonical brain networks, support the role of the visual word-form area in visuospatial attention, and investigate the heterogeneous organization of the frontoparietal control network.

SI Application of NeuroLang to the Resolution of Typical Queries on coordinate-based meta-analysis (CBMA) Databases

NeuroLang can be applied to the resolution of term-based and coactivation queries on CBMA databases, which are the most common types of meta-analyses conducted in the literature. A *term-based* query derives an activation pattern associated with a term of interest that relates to psychological concepts or cognitive processes. A *coactivation* query delineates a brain activation pattern that comprises spatially distant brain regions, putatively forming a large-scale functional network. Answering these queries is typically done using tools like Neurosynth¹ or BrainMap².

SI.1 From Term to Association Map

A core goal of meta-analyses is to identify brain regions that are preferentially activated in experiments studying a psychological or cognitive process, relative to the rest of the brain³. As an example, a meta-analysis of studies on *emotion* should find that the amygdala is one of the regions that is the most likely to be reported by this set of studies. In other words, the meta-analysis should find that there is an association between studies of emotion and neural response in the amygdala. Several models exist for computing probabilistic brain maps from a set of studies associated with a given term of interest (‘emotion’, in this example). All provide an estimate of the probability that a voxel gets reported by studies associated with the term. We explore and implement two of them in this example: multilevel kernel density analysis (MKDA) and activation likelihood estimation (ALE).

An important point should be made first, regarding the differences between approaches to combining the peak activation coordinates that are reported by neuroimaging studies. Peak activations are 3-dimensional points in the brain that lives in a standardized stereotactic coordinate system and that can be seen as a sparse representation of the statistical map obtained from an experimental study of neuroimaging signals. To account for the noise and variability of the *exact location* of these peaks, meta-analysis models often assume a neighborhood around peaks to be reported, or to have a probability of being reported. The definition of this spatial neighborhood is an *a priori* of the meta-analysis is sometimes called a spatial prior. This process, applied to the map of each study separately, is called *smoothing*⁴.

A simple but common spatial smoothing approach is to consider all voxels within a sphere centered at a reported peak location’s coordinates to also be reported by the study. This prior, used by MKDA, is illustrated on the left part of fig. S1. MKDA obtains one such binary pattern for each study and aggregates them as frequencies at which each voxel is reported by studies selected for a meta-analysis. In plain English, a voxel at location (x, y, z) is reported by a study s whenever s reports a

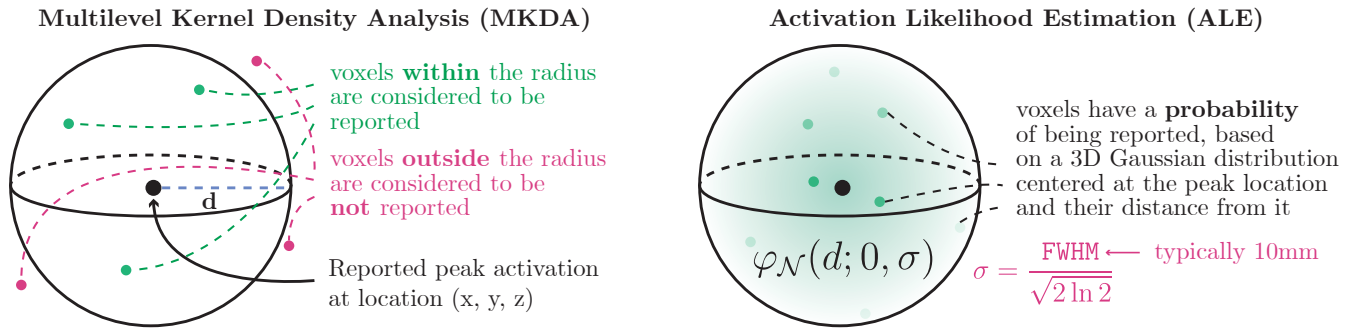


Figure S1. Illustration of spatial smoothing priors used by MKDA and ALE to address the uncertainty in the location of peak activation coordinates reported by neuroimaging studies.

peak activation within 10mm of that voxel. In NeuroLang, this sentence is encoded by the following `VoxelReported` rule

```
VoxelReported(x, y, z, s) :-
  PeakReported(x2, y2, z2, s) & Voxel(x, y, z)
  & d = euclidean_distance(x, y, z, x2, y2, z2) & d < 10
```

where the pairwise Euclidean distance between pairs of voxels is calculated using a built-in `euclidean_distance` function. The lower bound $d < 10$ defines a 10mm radius ball around each peak location⁵. An MKDA meta-analytic map is obtained by estimating the conditional probability that each voxel gets reported by studies associated with a given term of interest (e.g. ‘emotion’.) To compute estimates of these probabilities, we define the following NeuroLang rule, which gives each study an equal probability of being selected

```
SelectedStudy(s_1) : 1/N, ..., SelectedStudy(s_N) : 1/N :- T
```

The distribution of the program’s outputs can then be used to estimate the probability of each voxel being reported. Then, the result of solving a conditional probabilistic rule

```
ans(x, y, z, PROB) :-
  VoxelReported(x, y, z, s) & SelectedStudy(s)
  // TermAssociation(emotion, s) & SelectedStudy(s)
```

will contain a probability for each voxel of the brain, based on the frequency at which it is reported by studies within the database. The `//` operator represents a *probabilistic conditioning*.

Another approach to smoothing peak activations is to model their location’s uncertainty *probabilistically*. We give the example of the ALE algorithm⁶, illustrated on the right part of fig. S1. Instead of using a hard sphere, a three-dimensional Gaussian probability distribution is defined at each peak. This distribution provides a probability that depends on the Euclidean distance between each reported peak and brain voxel. The standard deviation of the Gaussian distribution is directly linked to the user-specified FWHM, as in the ALE algorithm. In NeuroLang, this is modeled using a rule that defines a `VoxelReported` table is as follows

```
VoxelReported(x, y, z, s) : p :-
  PeakReported(x2, y2, z2, s) & Voxel(x, y, z)
  & d = euclidean_distance(x, y, z, x2, y2, z2) & d < 2 * FWHM
  & p = gaussian(d ; 0, sigma) & FWHM = 10
  & sigma = FWHM / 2 * sqrt(2 * log(2))
```

Following the ALE model, the probability of activation for voxel (x, y, z) is calculated as the probability that any study in the meta-analysis report its activation. Assuming voxel reporting events to be independent, this means that the probability $\mathbf{P}[\exists s, \text{VoxelReported}(x, y, z, s) \wedge \text{TermAssociation}(\text{emotion}, s)]$ is equal to

$$1 - \prod_s (1 - \mathbf{P}[\text{VoxelReported}(x, y, z, s) \wedge \text{TermAssociation}(\text{emotion}, s)]) \quad (1)$$

where s is a study, as described in⁷. To estimate this probability, we use an *existentially-quantified* rule — that is, a rule that uses a predicate logic existential (\exists) quantifier — as follows

```

ans(x, y, z, PROB) :- exists(s;
  VoxelReported(x, y, z, s)
  & TermAssociation("emotion", s)
)

```

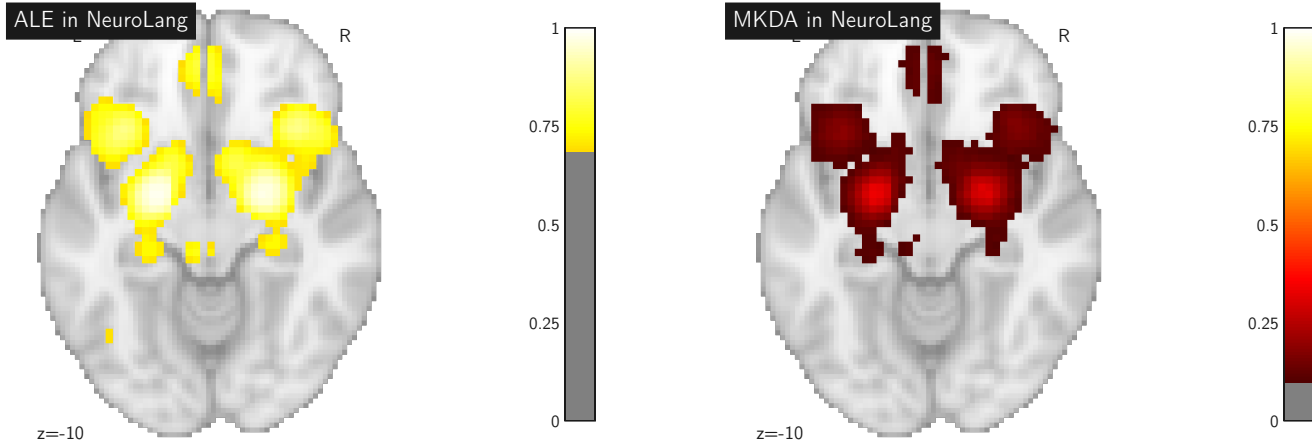


Figure S2. Probability maps for the term ‘emotion’, obtained from the MKDA and ALE NeuroLang programs. Values are shown before any statistical test is performed. The maps are thresholded to select only the top 1% probabilities.

In fig. S2, we compare the probability maps obtained from the MKDA and ALE NeuroLang programs. As expected, the association maps obtained with the two methods highlight a cluster in the amygdala, which is known to be part of the neural signature of emotion.

Sl.2 From Seed-Voxel to Coactivation Map

We now show how meta-analyses of voxel coactivations can be formulated in NeuroLang. These analyses are able to capture some of the most studied functional networks of the brain: the frontoparietal cognitive control network, dorsal attention network, and sensorimotor network. This was already shown to be possible on the manually curated BrainMap database⁸. We replicate these results, with NeuroLang, on the automatically extracted NeuroQuery CBMA database⁹. To solve coactivation queries in NeuroLang, we use the same program as in the previously described MKDA example, although the `TermAssociation` table plays no role now, as it is not needed for solving coactivation queries. We formulate the following conditional probabilistic rule

```

ans(x, y, z, PROB) :-
  VoxelReported(x, y, z, s) & SelectedStudy(s)
  // VoxelReported(x2, y2, z2, s) & SeedVoxel(x2, y2, z2)
  & SelectedStudy(s)

```

which asks, in plain English, ‘what is the probability that each voxel at MNI coordinates (x,y,z) gets reported by studies reporting the activation of some seed voxel at MNI coordinates (x_2,y_2,z_2) ’. This is solved for every voxel within 10mm of any peak reported at least once in the database, because the MKDA spatial smoothing prior is used. Note that this rule computes the coactivation maps for each seed voxel within the `SeedVoxel` table, meaning that the query needs to be solved only once to obtain several meta-analytical results. The `SeedVoxel` table contains the same coordinates as the ones used by Toro et al.⁸.

In fig. S3, we show the resulting coactivation map for each seed voxel. The first coactivation map, with a seed voxel in the left central sulcus, correctly highlights the sensorimotor network, with clear activation clusters in the precentral and postcentral gyri¹⁰. The second coactivation map, with a seed voxel in the ACC, correctly highlights the DMN, with activation clusters in the medial prefrontal cortex, posterior cingulate cortex, and nucleus accumbens. Finally, the third coactivation map, with a seed voxel in the left intraparietal sulcus, correctly highlights the dorsal attention network, with activation clusters in the supplementary motor area, anterior insula, frontal eye fields, dorsolateral prefrontal cortex, and ventral occipital cortex. These coactivation maps correctly highlight the functional network their respective seed voxel is expected to belong to, similar to what was reported by Toro et al.⁸.

To conclude, we have shown that typical term-based and coactivation-based queries on CBMA could be expressed and solved with NeuroLang.

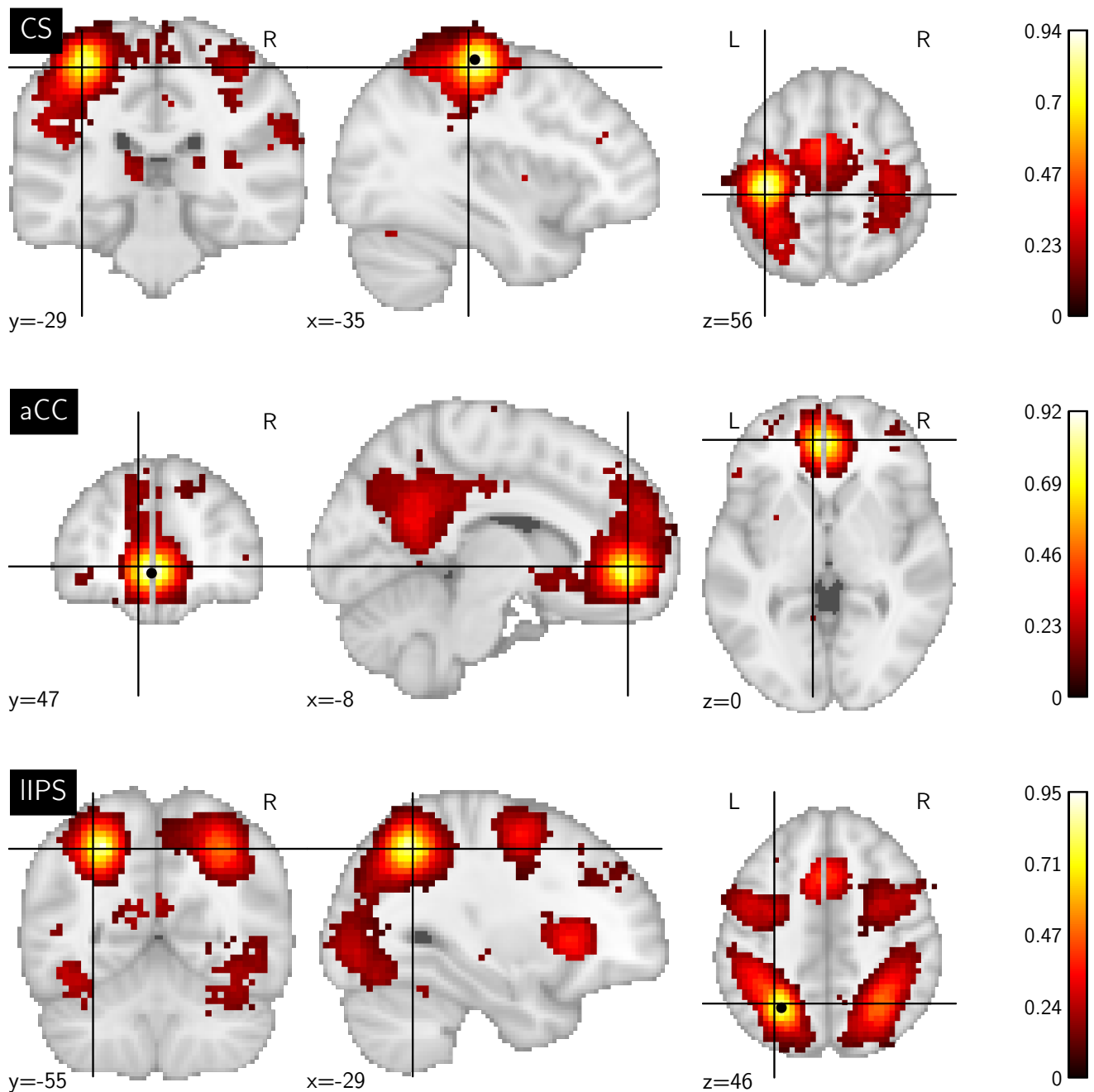


Figure S3. Coactivation maps obtained from seed voxels in the dorsal part of the left central sulcus ($x = -34, y = -26, z = 60$), the ACC ($x = -2, y = 46, z = -4$), and the left intraparietal sulcus ($x = -26, y = -58, z = 48$). The number of studies reporting activations within 10mm of each seed voxel was 971, 1812, and 1380 1380, 1812 and 971 for the central sulcus, the ACC, and the left intraparietal sulcus (respectively).

SII VWFA: Term Associations not Surviving Correction for Multiple Comparison

In table S1, we complement our topic-based study of the VWFA but with an analysis of raw terms within the Neurosynth database. The terms are indeed related to the topics found in our experiments: object and word recognition. However, none of these term associations survive an FDR-based correction for multiple comparisons. The calculation is run for all terms within Neurosynth's vocabulary. We conducted the same analysis on the NeuroQuery database, which contains a broader vocabulary and more term associations, and obtained similar results.

Term associated with studies reporting the VWFA and the ‘language’ network, but not reporting the frontoparietal attention network

Term	<i>p</i> -val (uncorrected)
l2	0.000773
bilinguals	0.001411
l1	0.003706
sublexical	0.006793
visual word recognition	0.006987

Term associated with studies reporting the VWFA and the fronto-parietal attention network, but not reporting the ‘language’ network

Term	<i>p</i> -val (uncorrected)
occipitotemporal cortex	0.000008
object recognition	0.000145
conflict detection	0.007809
spatial processing	0.007809
visual working memory	0.009181
visual word recognition	0.009684
word recognition	0.009684
ba37	0.009684
visual stream	0.014894
exner	0.024605
fonts	0.032104

Table S1. Term associations surviving a likelihood ratio test ($p < 0.05$).

References

1. Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C. & Wager, T. D. Large-scale automated synthesis of human functional neuroimaging data. *Nat Methods* **8**, 665–670, DOI: [10.1038/nmeth.1635](https://doi.org/10.1038/nmeth.1635) (2011).
2. Laird, A. R., Lancaster, J. L. & Fox, P. T. BrainMap: The social evolution of a human brain mapping database. *NI* **3**, 065–078, DOI: [10.1385/ni.3:1:065](https://doi.org/10.1385/ni.3:1:065) (2005).
3. Yarkoni, T., Poldrack, R. A., Van Essen, D. C. & Wager, T. D. Cognitive neuroscience 2.0: Building a cumulative science of human brain function. *Trends Cogn. Sci.* **14**, 489–496, DOI: [10.1016/j.tics.2010.08.004](https://doi.org/10.1016/j.tics.2010.08.004) (2010). Edition: 2010/09/29.
4. Samartsidis, P., Montagna, S., Johnson, T. D. & Nichols, T. E. The coordinate-based meta-analysis of neuroimaging data. *Stat. Sci.* **32**, 580–599, DOI: [10.1214/17-sts624](https://doi.org/10.1214/17-sts624) (2017).
5. Wager, T. D., Lindquist, M. & Kaplan, L. Meta-analysis of functional neuroimaging data: Current and future directions. *Soc. Cogn. Affect. Neur.* **2**, 150–158, DOI: [10.1093/scan/nsm015](https://doi.org/10.1093/scan/nsm015) (2007).
6. Eickhoff, S. B., Bzdok, D., Laird, A. R., Kurth, F. & Fox, P. T. Activation likelihood estimation meta-analysis revisited. *Neuroimage* **59**, 2349–2361, DOI: [10.1016/j.neuroimage.2011.09.017](https://doi.org/10.1016/j.neuroimage.2011.09.017) (2012).
7. Eickhoff, S. B. *et al.* Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: A random-effects approach based on empirical estimates of spatial uncertainty. *Hum. Brain Mapp.* **30**, 2907–2926, DOI: [10.1002/hbm.20718](https://doi.org/10.1002/hbm.20718) (2009). Publisher: John Wiley & Sons, Ltd.
8. Toro, R., Fox, P. & Paus, T. Functional coactivation map of the human brain. *Cereb. Cortex* **18**, 2553–2559, DOI: [10.1093/cercor/bhn014](https://doi.org/10.1093/cercor/bhn014) (2008).
9. Dockès, J. *et al.* NeuroQuery, comprehensive meta-analysis of human brain mapping. *eLife* **9**, e53385, DOI: [10.7554/elife.53385](https://doi.org/10.7554/elife.53385) (2020).
10. Chenji, S. *et al.* Investigating default mode and sensorimotor network connectivity in amyotrophic lateral sclerosis. *PLoS ONE* **11**, e0157443, DOI: [10.1371/journal.pone.0157443](https://doi.org/10.1371/journal.pone.0157443) (2016). Publisher: Public Library of Science.