

# Complex Coordinate-Based Meta-Analysis with Probabilistic Programming

Valentin Iovene, Gaston E Zanitti, Demian Wassermann

Inria, CEA, Université Paris-Saclay, Palaiseau, France  
 tgy@inria.fr, gaston.zanitti@inria.fr, demian.wassermann@inria.fr

## Abstract

With the growing number of published functional magnetic resonance imaging (fMRI) studies, meta-analysis databases and models have become an integral part of brain mapping research. Coordinate-based meta-analysis (CBMA) databases are built by extracting both coordinates of reported peak activations and term associations using natural language processing techniques from neuroimaging studies. Solving term-based queries on these databases makes it possible to obtain statistical maps of the brain related to specific cognitive processes. However, existing tools for analysing CBMA data are limited in their expressivity to propositional logic, restricting the variety of their queries. Moreover, with tools like Neurosynth, term-based queries on multiple terms often lead to power failure, because too few studies from the database contribute to the statistical estimations. We design a probabilistic domain-specific language (DSL) standing on Datalog and one of its probabilistic extensions, CP-Logic, for expressing and solving complex logic-based queries. We show how CBMA databases can be encoded as probabilistic programs. Using the joint distribution of their Bayesian network translation, we show that solutions of queries on these programs compute the right probability distributions of voxel activations. We explain how recent lifted query processing algorithms make it possible to scale to the size of large neuroimaging data, where knowledge compilation techniques fail to solve queries fast enough for practical applications. Finally, we introduce a method for relating studies to terms probabilistically, leading to better solutions for two-term conjunctive queries (CQs) on smaller databases. We demonstrate results for two-term CQs, both on simulated meta-analysis databases and on the widely used Neurosynth database.

## Introduction

The non-invasivity of functional magnetic resonance imaging (fMRI) led it to dominate brain mapping research since the early 1990s (Huettel, Song, and McCarthy 2008). In the past three decades, tens of thousands of published studies acquired and analysed fMRI signals, producing new understanding of the human brain and the cognitive function of its different components. Quickly, the idea of *meta-analysing* this ever-growing amount of neuroimaging studies flourished. By gathering and synthesising the findings of a large

corpora of neuroimaging studies, can we derive new knowledge about the brain’s mechanics? Can we study consensus within the cognitive neuroscience community? The lack of power of neuroimaging studies undermines the reproducibility of their findings (Poldrack et al. 2017; Botvinik-Nezer et al. 2020). Can we *build* consensus by aggregating results from several underpowered studies into more robust findings supported by past literature? Neuroimaging studies traditionally report peak activation coordinates in a standard stereotactic coordinate system. This makes it possible to compare them from one study to another. These coordinates result from statistical hypothesis testing and represent a condensed synthesis of which regions of the brain are reported as activated by a given study. Directly meta-analysing whole-brain unthresholded statistical maps (i.e. *image*-based meta-analysis) is known to yield statistically more powerful results (Salimi-Khorshidi et al. 2009); and the field is moving in that direction (Gorgolewski et al. 2015). However, most meta-analyses have resorted to coordinate-based meta-analysis (CBMA): the meta-analysis of these peak activation coordinates. In the past ten years, an ecosystem of CBMA databases and tools was brought to life (e.g. Laird, Lancaster, and Fox 2005; Yarkoni et al. 2011), becoming an integral part of brain mapping research. Automatic CBMA databases, like Neurosynth, extract both natural language processing features and peak activation coordinates from neuroimaging studies. These tools are used to derive activation patterns (e.g. Wager et al. 2013; Cole et al. 2012) or reveal meaningful cognitive processes through reverse inference (e.g. Smallwood and Schooler 2015; Seghier 2013; Chang et al. 2013; Andrews-Hanna, Smallwood, and Spreng 2014). With them, researchers can define more robust regions of interest supported by past literature.

Nonetheless, currently available tools are limited in the complexity of queries that they can express and solve on CBMA data. Neurosynth (Yarkoni et al. 2011) presents brain map for single-term queries. Although technically feasible, term-based conjunctive queries (CQs) lead to underpowered meta-analyses due to the small number of studies matching the queries. Methods for exploiting CBMA data have recently been proposed, but they concern themselves with either developing new procedures for thresholding statistical brain maps, or integrating spatial priors into probabilistic models by correlating nearby voxel activations (Montagna

et al. 2018). We look at neuroimaging meta-analysis from a different angle by improving upon existing CBMA literature through the development of a domain-specific language (DSL) that leverages past research on probabilistic logic programming languages and databases to *formulate and solve more expressive CBMA queries*. We believe that, with this approach, more could be wrung out of this type of data.

Recently, NeuroQuery (Dockès et al. 2020) produced meta-analyses using unstructured text-based queries. By encoding the relationship between terms in a vocabulary using a regularised linear model, NeuroQuery can produce brain maps for underrepresented terms (few studies exactly match the term). However, NeuroQuery’s queries are distinct from and harder to interpret than database queries, which have clear semantics. Moreover, producing a brain map from studies related to some term  $t_1$  and not related to some other term  $t_2$  is not possible because NeuroQuery cannot express logic-based queries. Finally, NeuroQuery is not a probabilistic model that can be plugged into a richer hierarchical model combining meta-analyses with heterogeneous modalities, such as neuroanatomical and ontological knowledge.

Since the 1970s, the computer science community has been working on extending logic programming languages (Roussel 1975; Abiteboul, Hull, and Vianu 1995) with probabilistic semantics to represent knowledge uncertainty inherent to real-world data (reviewed by De Raedt and Kimmig 2015). A wide variety of efficient approaches to answering questions (queries) from these programs were developed, alongside seminal theoretical understandings. Domain-specific languages are not new to the cognitive neuroscience community. The White Matter Query Language (Wassermann et al. 2013) was developed to help experts formally describe white matter tracts in a near-to-English syntax. To the best of our knowledge, applying these techniques to the formulation and resolution of logic-based queries on probabilistic CBMA databases has yet to be attempted. This approach could make it possible to formulate elaborate hypotheses on the brain’s function and structure and test them against past cognitive neuroscience literature.

Adopting a language-oriented programming approach, we use probabilistic logic programming languages to formulate and solve logic-based queries on CBMA databases. This work fits into a broader project to design a DSL, coined *NeuroLang*, for expressing and testing cognitive neuroscience hypotheses that combine meta-analysis, neuroanatomical and ontological knowledge. *The work presented here focuses on the probabilistic semantics of NeuroLang and its application to term-based CBMA queries.*

Contributions of this work are three-fold. First, we investigate the feasibility and technicalities of applying probabilistic logic programming to CBMA-based brain mapping. We propose a way to encode a CBMA database as a probabilistic logic program based on CP-Logic (Vennekens, Denecker, and Bruynooghe 2009), on which complex CBMA queries can be solved. We translate this program to an equivalent Bayesian network representation in order to show that correct answers to probabilistic queries can be derived from its factorised joint probability distribution. Second, we explain

how leveraging lifted query processing techniques (Braz, Amir, and Roth 2005; Dalvi and Suciu 2012) allows us to scale to the large size of neuroimaging data at the voxel level. Third, we propose a relaxed modeling of TFIDF features to better encode the relationship between terms and studies and show that fewer samples are needed to solve two-term CQs than traditional approaches, on simulated and real CBMA databases.

## Background

### Term-Based Queries on CBMA Databases

An example of term-based query formulated in plain English is: “for each region of the brain, what is the probability that studies associated with both terms *insula* and *speech* report its activation?”. The result of term-based queries are used in forward inference to obtain a map of the brain’s activated regions reported by studies matching the query.

A CBMA database of  $N$  studies with a fixed vocabulary of  $M$  terms can be represented as two matrices  $\mathbf{X} \in \mathbb{R}^{N,M}$  and  $\mathbf{Y} \in \{0, 1\}^{N,K}$ , where  $\mathbf{X}_{ij}$  is a TFIDF feature measured for term  $j$  in study  $i$  and  $\mathbf{Y}_{ik} = 1$  if voxel  $k$  is reported as activated in study  $i$ . In practice,  $\mathbf{Y}$  is a sparse matrix because only a small proportion of voxels are reported within a single study.

Forward inference brain maps are constructed from a probabilistic model where binary random variables  $A_k$  and  $T_j$  respectively model the activation of each voxel  $v_k$  and the association of studies to each term  $t_j$ .  $\mathbf{P}[A_k|T_j]$  is the probability that voxel  $k$  activates in studies conditioned on the studies being associated with term  $j$  and  $\mathbf{P}[A_k|T_{\text{insula}} \wedge T_{\text{speech}}]$  is the probability that voxel  $k$  activates in studies conditioned on studies being associated with both terms ‘insula’ and ‘speech’.<sup>1</sup>

Neurosynth (Yarkoni et al. 2011) associates terms to studies by applying a threshold  $\tau$  to TFIDF features  $\mathbf{X}$ . Forward inference maps are obtained by estimating, for each voxel  $k$ ,

$$\mathbf{P}[A_k|T_j] = \frac{\sum_{i=1}^N \mathbf{Y}_{ik} \mathbf{1}[\mathbf{X}_{ij} > \tau]}{\sum_{i=1}^N \mathbf{1}[\mathbf{X}_{ij} > \tau]} \quad (1)$$

Solving a query with a  $p$ -term conjunction,  $\varphi = T_1 \wedge \dots \wedge T_p$ , is done by estimating, for each voxel  $k$ ,

$$\mathbf{P}[A_k|\varphi] = \frac{\sum_{i=1}^N \mathbf{Y}_{ik} \mathbf{1}[\min(\mathbf{X}_{i1}, \dots, \mathbf{X}_{ip}) > \tau]}{\sum_{i=1}^N \mathbf{1}[\min(\mathbf{X}_{i1}, \dots, \mathbf{X}_{ip}) > \tau]} \quad (2)$$

As terms are added to this conjunction (and thus, complexity to the query), the term  $\mathbf{1}[\min(\mathbf{X}_{i1}, \dots, \mathbf{X}_{ip}) > \tau]$  goes to zero for an increasing number of studies. Rapidly, obtaining a meaningful brain map becomes infeasible due to statistically weak results. A different model that relaxes the hard thresholding of TFIDF features is proposed in the next sections. Note that, solving a *disjunction* of two terms is done by replacing min with max, thereby requiring that only one of the TFIDF features passes the threshold. The more terms are added, the larger the number of studies that are included in the estimation. In that case, statistical power is thus not an issue.

<sup>1</sup>We use  $\mathbf{P}[A_k|T_i, T_j]$  to denote  $\mathbf{P}[A_k = 1|T_i = 1, T_j = 1]$ .

## Probabilistic Logic Programming

Before diving into how probabilistic logic programming can be used to encode CBMA data, we give a brief introduction to those languages through the example of CP-Logic. We also define the syntactic restrictions of the subset of this language that we use in our DSL.

**CP-Logic** We use CP-Logic (Vennekens, Denecker, and Bruynooghe 2009) as an intermediate representation in the compilation of our DSL. In CP-Logic, programs contain rules (also called *CP-Events*) of the form

$$(h_1 : p_1 \vee \dots \vee h_n : p_n) \leftarrow \varphi \quad (3)$$

where  $h_i$  are head predicates,  $p_i$  are probabilities such that  $\sum_i p_i \leq 1$ , and the implication rule’s body (also called *antecedent*)  $\varphi$  is a first-order logic formula. All variables occurring in the head (also called *consequent*) of the rule must also occur in  $\varphi$ . Such rules are interpreted as ‘ $\varphi$  being true causes one of the atoms  $h_i$  to be true’. Which  $h_i$  becomes true is drawn from the probability distribution defined by probabilities  $p_i$ . CP-Logic programs define a probability distribution over the set of *possible worlds* (Sato 1995) associated with possible executions of the probabilistic program.

### Syntactic Restrictions and Probabilistic Databases

Only a *subset* of CP-Logic’s expressive syntax is necessary to encode a CBMA database and formulate term-based queries on it. In NeuroLang, two kinds of rules are allowed. Deterministic rules  $(h : 1) \leftarrow \varphi$ , where  $\varphi$  is a conjunction of predicates and  $h$  is a single head predicate that is true with probability 1 whenever  $\varphi$  is true. Probabilistic rules whose body is  $\top$  (always true). If the head of the rule contains a single predicate, it is a probabilistic fact. If it contains more than one head predicate, it is a probabilistic choice. Moreover, recursive rules such as  $(A(x) : 0.3) \leftarrow A(y) \wedge B(x)$  are not permitted in the program.

With these syntactic restrictions, probabilistic rules define relations in a probabilistic database. If a rule has more than one head predicate, its tuples are mutually exclusive and partition the space of possible worlds. Queries with mutually exclusive predicates are rewritten to be compatible with probabilistic tuple-independent databases. Deterministic rules of the program define *unions of conjunctive queries (UCQs)* on these relations. A UCQ  $Q(x)$  is defined by a disjunction  $CQ_1(x), \dots, CQ_n(x)$ , where  $CQ_i(x)$  are CQs which conjunct logic literals. One major theoretical result in the field of probabilistic databases is the *dichotomy theorem* (Dalvi and Suciu 2012). It classifies UCQs based on their complexity: those that can be solved in polynomial time and those that are #P-hard, in the size of the database. A set of rules analyses the syntax of a given UCQ  $Q(x)$  to derive an algebraic expression that solves  $\mathbf{P}[Q(x)]$ : the probability of  $Q(x)$  being true over all possible groundings of the database (i.e. possible worlds). This resolution strategy is called *lifted query processing*. Guarantees on the efficiency of query resolution is of particular interest in the context of neuroimaging’s high-dimensional space.

## Probabilistic CBMA Databases

We now describe how CBMA data and queries can be encoded as a CP-Logic program. We then show how this program can be translated to a Bayesian network. We use its factorised joint probability distribution to analytically derive the same solutions for term-based queries as Neurosynth. Finally, we describe our approach to solving queries on this program using lifted query processing strategies.

### Encoding a CBMA Database as a Probabilistic Logic Program

The program of fig. 1 encodes a CBMA database. The equiprobable choice on the SelectedStudy relation partitions the space of possible worlds such that each one corresponds to a particular study. VoxelReported and TermInStudy relations encode matrices  $\mathbf{Y}$  and  $\mathbf{X}$ . We write the program such that solving the query  $\mathbf{P}[\text{Activation}(v)|\varphi]$ , where  $\varphi$  conjuncts and/or disjuncts TermAssociation( $t_j$ ) atoms, produces the probabilistic model of term-based CBMA queries described in the Background section. For instance, when defining

$$\varphi = \text{TermAssociation}(\text{insula}) \wedge \text{TermAssociation}(\text{speech})$$

$\mathbf{P}[\text{Activation}(v)|\varphi]$  is equivalent to the query  $\mathbf{P}[A_k | T_{\text{speech}} \wedge T_{\text{insula}}]$  described previously. We show that in the next section.

### Equivalence With Term-Based Query Solutions

To justify the design of the program in fig. 1, we translate it to an equivalent Bayesian network representation using the algorithm proposed by Meert, Struyf, and Blockeel (2008). The resulting Bayesian network is depicted in fig. 2 using plate-notation. To simplify the notation, we use  $A_k$ ,  $T_n$  and  $T_m$  to denote random variables Activation( $v_k$ ), TermAssociation( $t_n$ ), and TermAssociation( $t_m$ ). From the joint probability distribution defined by the Bayesian network, it can be derived that

$$\mathbf{P}[A_k, T_n, T_m] \quad (4)$$

$$= \sum_{i=1}^N \mathbf{P}[c^{SS} = i] \mathbf{P}[c_{ki}^{\text{VR}} = 1] \mathbf{P}[c_{ni}^{\text{TIS}} = 1] \mathbf{P}[c_{mi}^{\text{TIS}} = 1] \quad (5)$$

$$= \frac{1}{N} \sum_{i=1}^N Y_{ik} \mathbf{1}[X_{in} > \tau] \mathbf{1}[X_{im} > \tau] \quad (6)$$

and, similarly, that

$$\mathbf{P}[T_n, T_m] = \sum_{i=1}^N \mathbf{P}[c^{SS} = i] \mathbf{P}[c_{ni}^{\text{TIS}} = 1] \mathbf{P}[c_{mi}^{\text{TIS}} = 1] \quad (7)$$

$$= \frac{1}{N} \sum_{i=1}^N \mathbf{1}[X_{in} > \tau] \mathbf{1}[X_{im} > \tau] \quad (8)$$

From these two joint probability distributions, the solution of the conditional query can be derived using that  $\mathbf{P}[A_k | T_n, T_m] = \frac{\mathbf{P}[A_k, T_n, T_m]}{\mathbf{P}[T_n, T_m]}$ , which gives the formula of eq. (2), for  $p = 2$ .

$$\begin{aligned}
& (\text{TermInStudy}(t_j, s_i) : 1) \leftarrow \top. & \forall i \in N, \forall j \in M, \mathbf{X}_{ij} > \tau \\
& (\text{VoxelReported}(v_k, s_i) : 1) \leftarrow \top. & \forall i \in N, \forall k \in K, \mathbf{Y}_{ik} = 1 \\
& \left( \bigvee_{i=1}^N \text{SelectedStudy}(s_i) : \frac{1}{N} \right) \leftarrow \top. \\
& (\text{TermAssociation}(t) : 1) \leftarrow \exists s (\text{TermInStudy}(t, s) \wedge \text{SelectedStudy}(s)). \\
& (\text{Activation}(v) : 1) \leftarrow \exists s (\text{VoxelReported}(v, s), \text{SelectedStudy}(s)).
\end{aligned}$$

Figure 1: CP-Logic program encoding a probabilistic CBMA database.  $\text{TermInStudy}(t, s)$  models the presence of term  $t$  in study  $s$ .  $\text{VoxelReported}(v, s)$  encodes whether voxel  $v$  was reported in study  $s$ . The large  $\text{SelectedStudy}$  equiprobable choice over studies makes each possible world correspond to a specific study.  $\text{Activation}(v)$  and  $\text{TermAssociation}(t)$  respectively model the activation of voxel  $v$  and the association with term  $t$ . The SUCC query  $\mathbf{P}[\text{Activation}(v)]$  gives the marginal probability of activation of voxels over all studies. The query  $\mathbf{P}[\text{Activation}(t)|\text{TermAssociation}(\text{insula})]$  results in a forward inference map for the term *insula*.

The same can be shown for disjunctive queries  $\mathbf{P}[A_k|T_n \vee T_m]$  by summing the results of 3 two-term CQs as follows

$$\begin{aligned}
\mathbf{P}[A_k|T_n \vee T_m] &= \mathbf{P}[A_k|T_n, T_m] + \\
&\mathbf{P}[A_k|\neg T_n, T_m] + \mathbf{P}[A_k|T_n, \neg T_m] \quad (9)
\end{aligned}$$

This confirms that the probabilistic program of fig. 1 is sound, as solving queries on the program leads to the statistical estimation described in the previous section.

### Solving Queries on Probabilistic CBMA Databases

We now explore query resolution techniques that scale to the size of large probabilistic CBMA databases. The estimation of a forward inference brain map for a two-term conjunction corresponds to the query

$$\mathbf{P}[\text{Activation}(v)|\text{TermAssociation}(t_i), \text{TermAssociation}(t_j)]$$

We solve this task by defining two CQs

$$\begin{aligned}
Q_1(v) &\leftarrow \frac{\text{Activation}(v), \text{TermAssociation}(t_i)}{\text{TermAssociation}(t_j)}, \\
Q_2 &\leftarrow \text{TermAssociation}(t_i), \text{TermAssociation}(t_j)
\end{aligned}$$

such that  $\frac{\mathbf{P}[Q_1(v)]}{\mathbf{P}[Q_2]}$  solves the initial query. The numerator corresponds to the joint probability of voxel activation and association to both terms. The denominator corresponds to the joint probability of association to both two terms.

### Knowledge-Compilation Approaches Do Not Scale to the Size of Neuroimaging Data

We implemented the program of fig. 1 in ProbLog2 (Dries et al. 2015). We observed that, when solving two-term CQs, grounding and compiling the program to sentential decision diagrams (SDDs) was impractical. Solving a two-term CQ takes more than 30 minutes on a recent laptop. This is due to the large number of voxels, terms and studies modeled in the program, leading to a large number of ground literals. To give perspective on the scale of CBMA and neuroimaging data, a brain is typically partitioned into a grid of about 230,000 2mm<sup>3</sup> voxels. On average, studies in the Neurosynth database report 3165

voxel activations. There are 14,371 studies and 3228 terms in the Neurosynth database. We also tried compiling our program manually to SDDs (Darwiche 2011). Despite our efforts, which did not include exploring recent tree-building strategies (Amarilli, Bourhis, and Senellart 2016), the resolution of queries was still too slow to be practical for real-world applications. Currently available CBMA tools are capable of solving single-term queries in seconds. Resolution of more complex queries should have a similar time complexity.

### Lifted Processing of UCQs on Probabilistic CBMA Databases

We leverage theoretical results which have identified classes of queries that lifted inference can solve in polynomial time. The dichotomy theorem (proven in Dalvi and Suciu 2012) provides a procedure for checking that UCQs are *liftable*. This theorem is convenient because it guarantees that any query such that the lifted processing rules apply is guaranteed to be solvable in PTIME. If the query is not liftable, we resort to knowledge compilation (KC)-based resolution techniques. Because the language does not have probabilistic clauses and prevents recursivity, we can use its deterministic rules to construct UCQs associated with a given probabilistic query  $\mathbf{P}[\psi(\mathbf{x})]$ , where  $\psi(\mathbf{x})$  is a conjunction of intensional, extensional or probabilistic literals. This lifted approach makes it possible to solve CQ in a few seconds. *Extensional query plans* (see 4.1 of Van den Broeck and Suciu 2017) are obtained and evaluated to solve queries using a custom Python relational algebra engine.

### Relating Terms and Studies Probabilistically

The hard thresholding  $1[x > \tau]$  of TFIDF features  $x$  presented in the Background section, and used by Neurosynth, misses studies that could be relevant to the resolution of queries. Because we are interested in solving more complex queries, in this section we explore a relaxation by introducing the soft-thresholding function

$$\omega(x; \alpha, \tau) := \sigma(\alpha(x - \tau)) \in [0, 1] \quad (10)$$

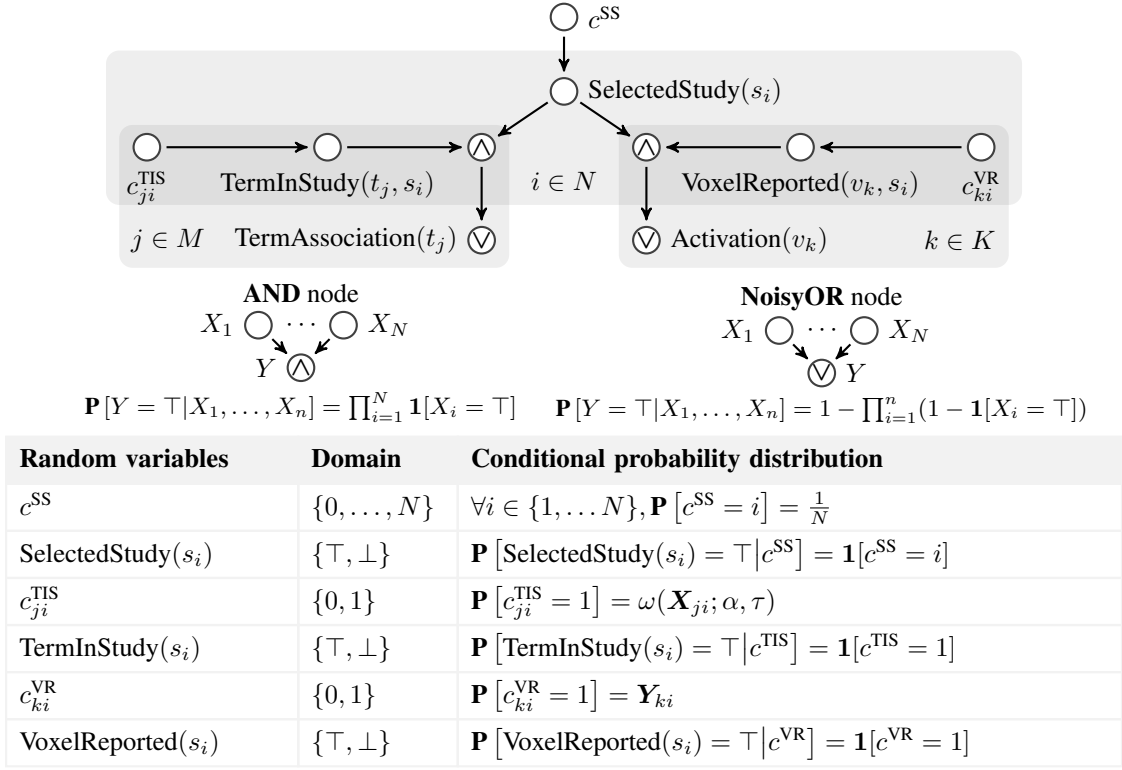


Figure 2: Plate-notation representation of the Bayesian network translated from the program proposed to encode a CBMA database. Each ground atom in the program (e.g.  $\text{TermInStudy}(t_2, s_{21})$ ) becomes a binary random variable with a deterministic conditional probability distribution (CPD). Specific AND nodes encode the conjunctions in the antecedent of the rules of the program. Choice random variables  $c^{SS}$ ,  $c_{ji}^{TIS}$ ,  $c_{ki}^{VR}$  represent probabilistic choices in the program.

where  $\sigma$  is the logistic function and  $\tau$  a threshold. As  $\alpha$  increases,  $\omega(x; \alpha, \tau)$  converges towards the hard-thresholding function  $\mathbf{1}[x > \tau]$ . With an appropriate  $\alpha$ , a larger proportion of studies is included in the calculation of  $\mathbf{P}[A_k | \varphi]$ , giving better estimates on small databases. For example, results of two-term CQ  $\mathbf{P}[A_k | T_1 \wedge T_2]$  and UCQ  $\mathbf{P}[A_k | T_1 \vee T_2]$  queries are estimated with

$$\mathbf{P}[A_k | T_1 \wedge T_2] = \frac{\sum_{i=1}^N \mathbf{Y}_{ik} \omega(\mathbf{X}_{i1}; \alpha, \tau) \omega(\mathbf{X}_{i2}; \alpha, \tau)}{\sum_{i=1}^N \omega(\mathbf{X}_{i1}; \alpha, \tau) \omega(\mathbf{X}_{i2}; \alpha, \tau)} \quad (11)$$

$$\mathbf{P}[A_k | T_1 \vee T_2] = \frac{\sum_{i=1}^N \mathbf{Y}_{ik} (1 - \prod_{j=1}^2 (1 - \omega(\mathbf{X}_{ij}; \alpha, \tau)))}{\sum_{i=1}^N (1 - \prod_{j=1}^2 (1 - \omega(\mathbf{X}_{ij}; \alpha, \tau)))} \quad (12)$$

More generally,  $\mathbf{P}[A_k | \varphi]$  can be estimated for first-order logic formulas  $\varphi$  that blend conjunctions and disjunctions of Boolean random variables  $T_j, j \in 1, \dots, M$ . For example,

if  $\varphi = (T_1 \vee T_2) \wedge (T_3 \vee T_4)$ , we have

$$\mathbf{P}[A_k | \varphi] = \frac{\sum_{i=1}^N \mathbf{Y}_{ik} f(\mathbf{X}_{i1}, \mathbf{X}_{i2}) f(\mathbf{X}_{i3}, \mathbf{X}_{i4})}{\sum_{i=1}^N f(\mathbf{X}_{i1}, \mathbf{X}_{i2}) f(\mathbf{X}_{i3}, \mathbf{X}_{i4})} \quad (13)$$

where  $f(x_1, x_2) = 1 - (1 - \omega(x_1; \alpha, \tau))(1 - \omega(x_2; \alpha, \tau))$ .

This modeling is implemented simply by integrating  $\omega(\mathbf{X}_{ij}; \alpha, \tau)$  as the probabilities of probabilistic facts  $\text{TermInStudy}(t_j, s_i)$  in the program of fig. 1.

## Experiments and Results

We compare our method with Neurosynth's on simulated CBMA databases sampled from a generative model and on the Neurosynth database. Using both models, we solve 55 different two-term CQs  $\mathbf{P}[A_k | T_i \wedge T_j]$ .

**Gain of Statistical Power when Solving Two-Term CQs on Smaller Simulated CBMA Databases** We evaluate our method on simulated small CBMA databases obtained by sampling from the generative model of fig. 3. This generative model provides the ground truth of which voxels activate in studies matching a given query of interest. This binary classification setting makes it possible to compare mod-

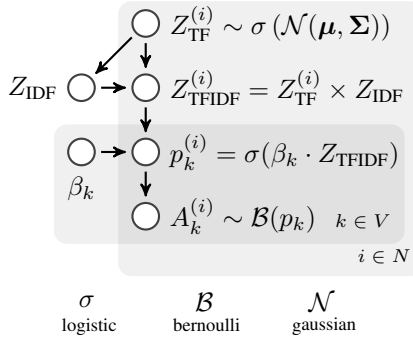


Figure 3: Model for generating CBMA databases of size  $N$ .  $Z_{\text{TF}}^{(i)}$  models term frequencies in study  $i$  and follows a logistic-normal distribution.  $Z_{\text{IDF}}$  computes inverse document frequencies from  $\{Z_{\text{TF}}^{(i)}\}_{i \in N}$ .  $p_k$  is the probability of activation of voxel  $v_k$ . Vectors  $\beta_k$  are obtained from a rejection sampling scheme that controls the proportion of voxels that activate when the query is verified.  $Z_{\text{IDF}}$ ,  $\mu$  and  $\Sigma$  are estimated from 4168 scrapped PubMed abstracts.

els by measuring their ability to identify true voxel activations for multiple sample sizes. We experimented with multiple numbers of voxels ( $K \in [100, 1000]$ ). Preliminary results showed that varying the number of voxels in this range does not alter the results. We report results for  $K = 1000$  voxels, of which 5% are activated in studies matching the query. Predicted voxel activations are obtained by thresholding  $p$ -values computed from each model’s estimation of  $\mathbf{P}[A_k|\varphi]$  using a  $G$ -test of independence. We use a  $p$ -value threshold of 0.01 and a Bonferroni correction for multiple comparisons. Simulation results for two-term CQs are presented in fig. 4, where we compare our model’s and Neurosynth’s  $F_1$  scores across 55 two-term CQs. These queries correspond to all two-term combinations out of 11 terms (depicted on the  $y$ -axis of the bottom plot of fig. 4) associated with a sufficiently large number of studies within the Neurosynth database to produce meaningful forward inference map. The  $F_1$  score measures the performance of a binary classifier by combining its precision and recall into a single metric. We see the advantage of our approach over Neurosynth’s for smaller generated samples where activations related to the query can be identified more reliably (higher  $F_1$  scores). Multiple values of  $\alpha$  in the range  $[100, 1000]$  were tried during our experiments. However when  $\alpha$  is too small or too large, the model tends to include either too many (and irrelevant) or too few studies in the estimation. When  $\alpha$  tends to 0,  $\omega$  becomes equivalent to Neurosynth’s hard thresholding. We found that a sweet spot for  $\alpha$  was around 300 and report results for that value. Drawing the sigmoid curve for  $\alpha = 300$  confirms that this transformation of TFIDF features is adequate because it maintains Neurosynth’s hard thresholding’s property of giving a 0 or 1 probability to the lowest and highest TFIDF features (respectively).

The proposed approach did not show an advantage over Neurosynth for solving two-term disjunctive queries. This

is expected, as such queries do not reduce the number of studies incorporated in the estimation of  $\mathbf{P}[A_k|T_i \vee T_j]$ , as explained in the Background section.

**Gain of Activation Consistency on a Real CBMA Database** We evaluate our method on the Neurosynth CBMA database. Because we don’t have a ground truth of which voxels activate for a given query, we resort to comparing models based on the *consistency* of their predicted activations over many random sub-samples of the Neurosynth database.

From predicted activation maps of  $K$  voxels obtained from  $M$  sub-samples of a CBMA database, the *consistency* for a two-term conjunctive formula  $\varphi$  is computed as

$$C_\varphi := \frac{1}{K} \sum_{k=1}^K \left( 1 - 2 \times \left| \frac{1}{M} \sum_{m=1}^M \hat{y}_{mk}^\varphi - \frac{1}{2} \right| \right) \quad (14)$$

where  $\hat{y}_{mk}^\varphi = 1$  if voxel  $k$  is predicted to be activated in sub-sample  $m$  when formula  $\varphi$  is true. The closer to one, the closer the average activation is to 0 or 1, which indicates a higher consistency across sub-samples. The closer to zero, the closer the average activation is to 0.5 which indicates that the predicted activations are highly variable across samples.

Results are reported in fig. 5, where the distribution of consistencies, across the same 55 CQs as in the previous experiment, are shown for multiple sample sizes. For the largest sample sizes, consistency scores are closer to 1 with our method than with Neurosynth’s. For a sample size of 2395 (chosen on a logarithmic scale), the average consistency of our method was 0.48 while Neurosynth’s was 0.4 (+20%) across samples and queries. For a sample size of 3856, we notice a 10% improvement.

We did not experiment with larger sample sizes due to the computational cost of running the experiment on many Neurosynth subsamples for all CQs. Also, we were mainly interested in whether our approach would be more consistent for smaller sample sizes. We observed that the consistency between Neurosynth and our approach was similar when both models were estimated on the entire database. This means that the proposed approach is more consistent on smaller sample sizes but equivalently consistent on larger sample sizes. The NeuroLang program implementing this model is available at [https://github.com/NeuroLang/NeuroLang/tree/master/examples/plot\\_neurosynth\\_relaxed\\_tfidf.py](https://github.com/NeuroLang/NeuroLang/tree/master/examples/plot_neurosynth_relaxed_tfidf.py).

## Discussion

This work fits into a broader approach to design a domain-specific language for expressing logic-based cognitive neuroscience hypotheses that combine neuroimaging data, neuroanatomical probabilistic maps, ontologies and meta-analysis databases to produce fine-grained brain maps supported by past literature and heterogeneous data.

The number of voxels ( $K = 1000$ ) used in the simulation experiments is orders of magnitude lower than on the typical whole-brain neuroimaging setting, where  $K \simeq \mathcal{O}(10^5)$ . We chose to lower the dimension in the simulation setting for computational practicality purposes. However, we believe that maintaining the same proportion of reported activations

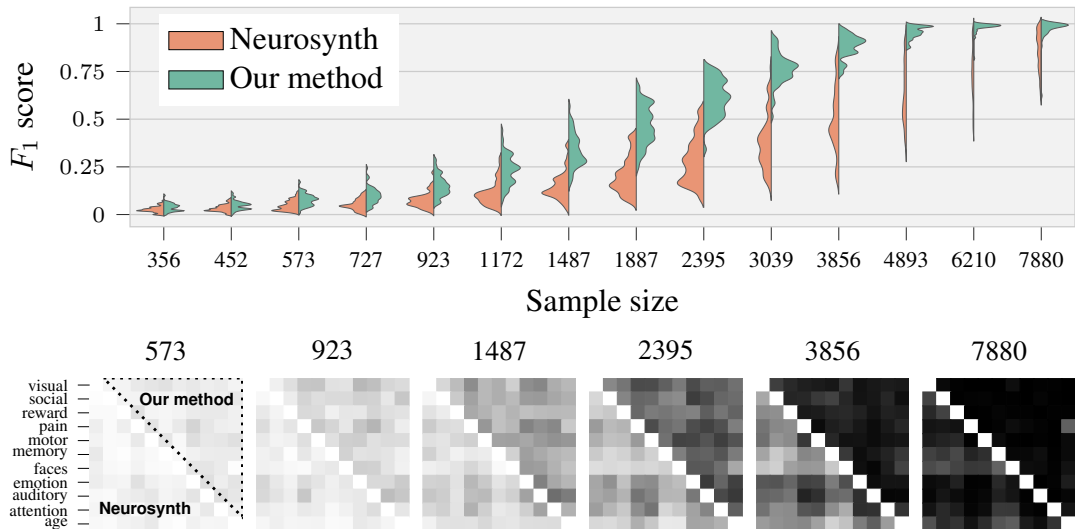


Figure 4: Comparison of Neurosynth’s and our method’s  $F_1$  scores across 55 two-term CQs on simulated CBMA databases of varying sample sizes. For each sample size, 100 random sub-samples were used. Above,  $F_1$  score distributions on all queries are compared across sample sizes. Below,  $F_1$  score matrices (white is 0, black is 1) are compared across sample sizes. The upper triangular contains scores of our method and the lower triangular contains scores of Neurosynth. The threshold  $\tau = 0.1$  is used in both models. The value  $\alpha = 300$  was empirically chosen. Varying  $\alpha$  near this value does not change the results noticeably. Sample sizes were taken on a logarithmic scale.

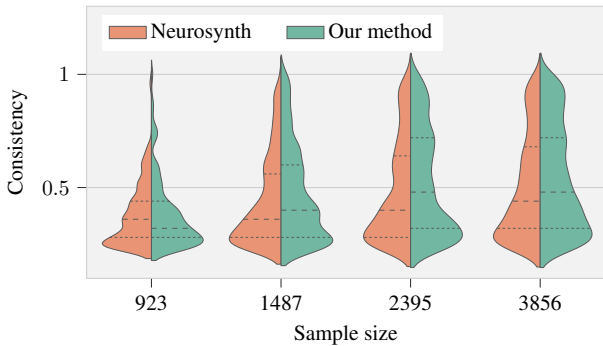


Figure 5: Comparison of both models’ distributions of voxel activation consistency across 1000 sub-samples of the Neurosynth’s database, for 55 two-term CQs and for multiple sample sizes. As the sample size increases, our method finds more consistent activations than Neurosynth.

as in the Neurosynth database was enough to confirm our approach on simulations before applying it to real data.

The flexible syntax of logic-based languages allows to express various kinds of queries.  $\mathbf{P}[\text{TermAssociation}(t) | \varphi_{\text{insula}} \vee \varphi_{\text{fMRI}}]$  queries for terms that are most probably associated with a given pattern of activation, where  $\varphi_{\text{insula}}$  is a conjunction of logic predicates  $\text{Activation}(v_k)$  whose probabilities come from a neuroanatomical probabilistic map of the insula, and where  $\varphi_{\text{fMRI}}$  is also a conjunction of predicates  $\text{Activation}(v_k)$  whose voxels  $v_k$  are based on neuroimaging data coming from a custom fMRI study.

The current version of NeuroLang is limited in what it can model, mainly due to the syntactic limitations on programs and queries that we had to make in order to use lifted processing strategies that scale to the size of CBMA data. In cognitive neuroscience, there is interest in using spatial priors to give the incentive to nearby voxels to co-activate (Kong et al. 2018). Spatial priors could be formulated as recursive probabilistic rules, such as  $\text{Activation}(v_1) : f(d) \leftarrow \text{Activation}(v_2), d = \text{euclidean}(v_1, v_2)$ , where  $d$  is the Euclidean distance measure between two regions of the brain, and where  $f$  maps  $d$  to a proper probability in  $[0, 1]$ . The resolution of such queries remains a challenge both in terms of methodology and tractability. Future progress in the field of probabilistic programming languages could open the door to other queries of interest to the cognitive neuroscience community.

## Conclusion

This work is a step towards incorporating complex meta-analyses in brain mapping models. We encode a CBMA database in a probabilistic logic program on which general logic-based queries can be solved. Leveraging efficient query resolution strategies on probabilistic databases, we are able to scale to the size of neuroimaging data. We experimented with a new method for solving two-term CQs using TFIDF features more efficiently than the hard thresholding scheme used by Neurosynth. This is promising but further investigation should be conducted to know whether this method extends to queries that conjunct more terms or queries that blend conjunctions and disjunctions. The proposed method requires the same computational power as Neurosynth.

## Acknowledgments

This work was funded by the ERC-2017-STG NeuroLang grant. We thank the AAAI-2021 reviewers and the senior program committee member for the quality and rigor of their comments on this work. We also thank the organisers of the conference for orchestrating the review of such a large number of submissions. We are grateful to some of our colleagues who shared insights on our work: Antonia Machlouzarides-Shalit, Hicham Janati, and Louis Rouillard-Odera.

## References

- Abiteboul, S.; Hull, R.; and Vianu, V. 1995. *Foundations of databases*. Reading, Mass: Addison-Wesley. ISBN 978-0-201-53771-0.
- Amarilli, A.; Bourhis, P.; and Senellart, P. 2016. Tractable Lineages on Treelike Instances: Limits and Extensions. In *PODS (Principles of Database Systems)*, 355–370. San Francisco, United States. URL <https://hal-imt.archives-ouvertes.fr/hal-01336514>.
- Andrews-Hanna, J. R.; Smallwood, J.; and Spreng, R. N. 2014. The default network and self-generated thought: component processes, dynamic control, and clinical relevance: The brain’s default network. *Annals of the New York Academy of Sciences* 1316(1): 29–52.
- Botvinik-Nezer, R.; Holzmeister, F.; Camerer, C. F.; Dreber, A.; Huber, J.; Johannesson, M.; Kirchler, M.; Iwanir, R.; Mumford, J. A.; Adcock, R. A.; Avesani, P.; Baczkowski, B. M.; Bajracharya, A.; Bakst, L.; Ball, S.; Barilari, M.; Bault, N.; Beaton, D.; Beitner, J.; Benoit, R. G.; Berkers, R. M. W. J.; Bhanji, J. P.; Biswal, B. B.; Bobadilla-Suarez, S.; Bortolini, T.; Bottenhorn, K. L.; Bowring, A.; Braem, S.; Brooks, H. R.; Brudner, E. G.; Calderon, C. B.; Camilleri, J. A.; Castrellon, J. J.; Cecchetti, L.; Cieslik, E. C.; Cole, Z. J.; Collignon, O.; Cox, R. W.; Cunningham, W. A.; Czoschke, S.; Dadi, K.; Davis, C. P.; Luca, A. D.; Delgado, M. R.; Demetriou, L.; Dennison, J. B.; Di, X.; Dickie, E. W.; Dobryakova, E.; Donnat, C. L.; Dukart, J.; Duncan, N. W.; Durnez, J.; Eed, A.; Eickhoff, S. B.; Erhart, A.; Fontanesi, L.; Fricke, G. M.; Fu, S.; Galván, A.; Gau, R.; Genon, S.; Glatard, T.; Glerean, E.; Goeman, J. J.; Golowin, S. A. E.; González-García, C.; Gorgolewski, K. J.; Grady, C. L.; Green, M. A.; Guassi Moreira, J. F.; Guest, O.; Hakimi, S.; Hamilton, J. P.; Hancock, R.; Handjaras, G.; Harry, B. B.; Hawco, C.; Herholz, P.; Herman, G.; Heunis, S.; Hoffstaedter, F.; Hogeveen, J.; Holmes, S.; Hu, C.-P.; Huettel, S. A.; Hughes, M. E.; Iacovella, V.; Iordan, A. D.; Isager, P. M.; Isik, A. I.; Jahn, A.; Johnson, M. R.; Johnstone, T.; Joseph, M. J. E.; Juliano, A. C.; Kable, J. W.; Kassinopoulos, M.; Koba, C.; Kong, X.-Z.; Koscik, T. R.; Kucukboyaci, N. E.; Kuhl, B. A.; Kupek, S.; Laird, A. R.; Lamm, C.; Langner, R.; Lauharatanahirun, N.; Lee, H.; Lee, S.; Leemans, A.; Leo, A.; Lesage, E.; Li, F.; Li, M. Y. C.; Lim, P. C.; Lintz, E. N.; Liphardt, S. W.; Losecaat Vermeer, A. B.; Love, B. C.; Mack, M. L.; Malpica, N.; Marins, T.; Maumet, C.; McDonald, K.; McGuire, J. T.; Melero, H.; Méndez Leal, A. S.; Meyer, B.; Meyer, K. N.; Mihai, G.; Mitsis, G. D.; Moll, J.; Nielson, D. M.; Nilsson, G.; Notter, M. P.; Olivetti, E.; Onicas, A. I.; Papale, P.; Patil, K. R.; Peelle, J. E.; Pérez, A.; Pischella, D.; Poline, J.-B.; Prys-tauka, Y.; Ray, S.; Reuter-Lorenz, P. A.; Reynolds, R. C.; Ricciardi, E.; Rieck, J. R.; Rodriguez-Thompson, A. M.; Romy, A.; Salo, T.; Samanez-Larkin, G. R.; Sanz-Morales, E.; Schlichting, M. L.; Schultz, D. H.; Shen, Q.; Sheridan, M. A.; Silvers, J. A.; Skagerlund, K.; Smith, A.; Smith, D. V.; Sokol-Hessner, P.; Steinkamp, S. R.; Tashjian, S. M.; Thirion, B.; Thorp, J. N.; Tinghög, G.; Tisdall, L.; Tompson, S. H.; Toro-Serey, C.; Torre Tresols, J. J.; Tozzi, L.; Truong, V.; Turella, L.; van ‘t Veer, A. E.; Verguts, T.; Vettel, J. M.; Vijayarajah, S.; Vo, K.; Wall, M. B.; Weeda, W. D.; Weis, S.; White, D. J.; Wisniewski, D.; Xifra-Porxas, A.; Yearling, E. A.; Yoon, S.; Yuan, R.; Yuen, K. S. L.; Zhang, L.; Zhang, X.; Zosky, J. E.; Nichols, T. E.; Poldrack, R. A.; and Schonberg, T. 2020. Variability in the analysis of a single neuroimaging dataset by many teams. *Nature* 582(7810): 84–88. ISSN 1476-4687. doi:10.1038/s41586-020-2314-9. URL <https://doi.org/10.1038/s41586-020-2314-9>.
- Braz, R. d. S.; Amir, E.; and Roth, D. 2005. Lifted First-Order Probabilistic Inference. In *IJCAI-05, Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence, Edinburgh, Scotland, UK, July 30 - August 5, 2005*, 1319–1325. Professional Book Center.
- Chang, L. J.; Yarkoni, T.; Khaw, M. W.; and Sanfey, A. G. 2013. Decoding the Role of the Insula in Human Cognition: Functional Parcellation and Large-Scale Reverse Inference. *Cerebral Cortex* 23(3): 739–749.
- Cole, M. W.; Yarkoni, T.; Repovs, G.; Anticevic, A.; and Braver, T. S. 2012. Global Connectivity of Prefrontal Cortex Predicts Cognitive Control and Intelligence. *Journal of Neuroscience* 32(26): 8988–8999.
- Dalvi, N.; and Suciu, D. 2012. The dichotomy of probabilistic inference for unions of conjunctive queries. *Journal of the ACM* 59(6): 1–87. Number: 6.
- Darwiche, A. 2011. SDD: A New Canonical Representation of Propositional Knowledge Bases. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume Two, IJCAI’11*, 819–826. AAAI Press.
- De Raedt, L.; and Kimmig, A. 2015. Probabilistic (logic) programming concepts. *Machine Learning* 100(1): 5–47.
- Dockès, J.; Poldrack, R. A.; Primet, R.; Gözükan, H.; Yarkoni, T.; Suchanek, F.; Thirion, B.; and Varoquaux, G. 2020. NeuroQuery, comprehensive meta-analysis of human brain mapping. *eLife* 9: e53385.
- Dries, A.; Kimmig, A.; Meert, W.; Renkens, J.; Van den Broeck, G.; Vlasselaer, J.; and De Raedt, L. 2015. ProbLog2: Probabilistic Logic Programming. In *Machine Learning and Knowledge Discovery in Databases*, volume 9286, 312–315. Cham: Springer International Publishing. Series Title: Lecture Notes in Computer Science.
- Gorgolewski, K. J.; Varoquaux, G.; Rivera, G.; Schwarz, Y.; Ghosh, S. S.; Maumet, C.; Sochat, V. V.; Nichols, T. E.; Poldrack, R. A.; Poline, J.-B.; Yarkoni, T.; and Margulies, D. S.



2015. NeuroVault.org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain. *Frontiers in Neuroinformatics* 9.
- Huettel, S. A.; Song, A. W.; and McCarthy, G. 2008. *Functional magnetic resonance imaging*. Sunderland, Mass: Sinauer Associates, 2nd ed edition.
- Kong, R.; Li, J.; Orban, C.; Sabuncu, M. R.; Liu, H.; Schaefer, A.; Sun, N.; Zuo, X.-N.; Holmes, A. J.; Eickhoff, S. B.; and Yeo, B. T. T. 2018. Spatial Topography of Individual-Specific Cortical Networks Predicts Human Cognition, Personality, and Emotion 19.
- Laird, A. R.; Lancaster, J. L.; and Fox, P. T. 2005. BrainMap: The Social Evolution of a Human Brain Mapping Database. *Neuroinformatics* 3(1): 065–078.
- Meert, W.; Struyf, J.; and Blockeel, H. 2008. Learning Ground CP-Logic Theories by Leveraging Bayesian Network Learning Techniques. *Fundam. Inform.* 89: 131–160.
- Montagna, S.; Wager, T.; Barrett, L. F.; Johnson, T. D.; and Nichols, T. E. 2018. Spatial Bayesian latent factor regression modeling of coordinate-based meta-analysis data: Spatial Bayesian Latent Factor Regression Modeling of CBMA Data. *Biometrics* 74(1): 342–353.
- Poldrack, R. A.; Baker, C. I.; Durnez, J.; Gorgolewski, K. J.; Matthews, P. M.; Munafò, M. R.; Nichols, T. E.; Poline, J.-B.; Vul, E.; and Yarkoni, T. 2017. Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience* 18(2): 115–126. ISSN 1471-0048. doi:10.1038/nrn.2016.167. URL <https://doi.org/10.1038/nrn.2016.167>.
- Roussel, P. 1975. *PROLOG: Manuel de Reference et d'Utilisation*. Université d'Aix-Marseille II.
- Salimi-Khorshidi, G.; Smith, S. M.; Keltner, J. R.; Wager, T. D.; and Nichols, T. E. 2009. Meta-analysis of neuroimaging data: a comparison of image-based and coordinate-based pooling of studies. *NeuroImage* 45(3): 810–823.
- Sato, T. 1995. A Statistical Learning Method for Logic Programs with Distribution Semantics. In *ICLP*.
- Seghier, M. L. 2013. The Angular Gyrus: Multiple Functions and Multiple Subdivisions. *The Neuroscientist* 19(1): 43–61.
- Smallwood, J.; and Schooler, J. W. 2015. The Science of Mind Wandering: Empirically Navigating the Stream of Consciousness. *Annual Review of Psychology* 66(1): 487–518.
- Van den Broeck, G.; and Suci, D. 2017. Query Processing on Probabilistic Data: A Survey. *Foundations and Trends® in Databases* 7(3-4): 197–341. Number: 3-4.
- Vennekens, J.; Denecker, M.; and Bruynooghe, M. 2009. CP-logic: A language of causal probabilistic events and its relation to logic programming. *Theory and Practice of Logic Programming* 9(3): 245–308.
- Wager, T. D.; Atlas, L. Y.; Lindquist, M. A.; Roy, M.; Woo, C.-W.; and Kross, E. 2013. An fMRI-Based Neurologic Signature of Physical Pain. *New England Journal of Medicine* 368(15): 1388–1397.
- Wassermann, D.; Makris, N.; Rathi, Y.; Shenton, M.; Kikinis, R.; Kubicki, M.; and Westin, C.-F. 2013. On Describing Human White Matter Anatomy: The White Matter Query Language. In *Advanced Information Systems Engineering*, volume 7908, 647–654. Berlin, Heidelberg: Springer Berlin Heidelberg. Series Title: Lecture Notes in Computer Science.
- Yarkoni, T.; Poldrack, R. A.; Nichols, T. E.; Van Essen, D. C.; and Wager, T. D. 2011. Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods* 8(8): 665–670.