https://doi.org/10.36315/2025inpact145
Psychological Applications and Trends 2025

CAN COGNITIVE FUNCTIONS BE INFERRED FROM NEUROIMAGING DATA? A REVERSE INFERENCE META-ANALYSIS OF THEORY-OF-MIND TASKS

Donato Liloia, & Tommaso Costa
Department of Psychology, University of Turin (Italy)

Abstract

Cognitive neuroscience research using functional magnetic resonance imaging (fMRI) has predominantly focused on localizing patterns of neural activity associated with human cognitive functions. This approach, known as *forward inference*, has been pivotal in pinpointing brain areas engaged during specific cognitive tasks and testing hypotheses about brain-behavior relationships. In contrast, the use of reasoning from brain activation to cognitive functions, known as *reverse inference*, has been considered more informative because it allows researchers to interpret neural activity patterns to make inferences about the cognitive domain likely at play. Crucially, *reverse inference* considers how selectively the area is activated by the cognitive function under investigation, which is particularly important given the multifunctional nature of many cortical and subcortical areas. Nevertheless, the practical application of *reverse inference* in fMRI research remains methodologically challenging. Here, we performed a meta-analytic *reverse inference* analysis of brain activations related to Theory-of-Mind (ToM) tasks to evaluate whether this approach can effectively identify selective brain areas recruited for this critical human cognitive function. Leveraging data from the BrainMap database, we analyzed 223 published fMRI experiments involving ToM tasks (1069 healthy participants and 1526 activation foci) and compared these findings to fMRI data from other tasks stored in the BrainMap database (110 distinct cognitive tasks, 8154 published experiments, 127112 healthy participants, and 66649 activation foci). To achieve this, we applied Bayes fACtor mOdeliNg, a novel Bayesian-based, data-driven, hypothesis-free method that provides posterior probability distributions for the evidence of selectivity with respect to a given mental process. We found that several brain areas commonly recruited in ToM tasks (e.g., bilateral inferior frontal gyri, superior temporal cortices, and posterior cingulate cortex) show a low level of selectivity ($P < 50\%$), indicating their involvement across multiple cognitive domains. The results also revealed a small, organized set of highly selective areas ($P > 90\%$; e.g., bilateral superior frontal gyri, inferior temporal gyri, right precuneus, and anterior cingulate cortex) that map the cognitive function of ToM. These results provide a more refined and nuanced approach to understanding the neural basis of cognition, offering valuable insights for the development of formal cognitive ontologies and the refinement of brain-cognition models.

Keywords: *Neuroimaging, fMRI, cognitive ontology, Bayesian modeling, social cognition.*

1. Introduction

Over the past three decades, neuroimaging techniques such as task-based functional magnetic resonance imaging (fMRI) have enabled the measurement of local brain activity in response to cognitive tasks performed during scanning. These data allow researchers to investigate the involvement of brain regions during cognitive functions and test hypotheses regarding brain-behavior relationships (Westbrook, 2021).

From a methodological point of view, such findings support a task-to-activation estimation, commonly referred to as *forward inference* (Henson, 2006). This reasoning approach is widely applied in task-based fMRI research and has been instrumental in identifying neural substrates underlying cognitive functions. However, a major limitation of *forward inference* is its lack of selectivity (Costa et al., 2021; Poldrack, 2006), which prevents determining whether a given brain region is selectively engaged in a specific function. Moreover, the involvement of many brain areas in multiple cognitive tasks (Cauda et al., 2012) further complicates task-to-activation estimations, limiting the contribution of fMRI findings to refining brain-cognition models.

Starting from the seminal work of Poldrack (2006), a complementary reasoning approach has been introduced to assess how selectively a brain region is activated by a given cognitive function: *reverse inference*. This approach may infer cognitive functions from observed activations, providing a framework for linking brain activity patterns to specific cognitive processes.

The application of *reverse inference* in fMRI research has been explored and extensively debated (Cauda et al., 2020; Costa et al., 2021; Poldrack, 2008). While many studies underscore its importance for fMRI research, several challenges persist. One major limitation is the absence of a comprehensive formal cognitive ontology, which restricts the ability to accurately infer cognitive functions from neuroimaging data (Poldrack et al., 2011). Additionally, the vast and continuously growing body of literature complicates efforts to establish strong claims about selective brain-function mappings. However, open-access repositories such as the BrainMap database (Fox & Lancaster, 2002) may mitigate this issue by consolidating a broad sample of peer-reviewed experiments, facilitating large-scale meta-analytic approaches.

Bayesian statistical models have been proposed since the earliest theorization of *reverse inference* in fMRI as a promising avenue for enhancing its reliability in cognitive neuroscience. However, only in recent years has the Bayesian statistics been implemented in a user-friendly and open-access tool called Bayes fACtor mOdeliNg (BACON) (Costa et al., 2021). This meta-analytic approach assesses the likelihood that a given activation pattern corresponds to a targeted cognitive function in a whole-brain, voxelwise, data-driven and hypothesis-free manner.

2. Objectives

This study conducted a *reverse inference* analysis of fMRI-based brain activations associated with Theory-of-Mind (ToM) tasks to determine whether BACON can identify brain regions with a high posterior probability of being selectively involved in this crucial cognitive function.

ToM - the ability to infer and predict the intentions, thoughts, and beliefs of others (Premack & Woodruff, 1978) - was chosen as a case study. This selection is motivated by the well-established identification of a "core brain network" for ToM in fMRI research, including the anterior and posterior cingulate cortex, precuneus, inferior, middle, and superior temporal gyri, angular gyrus, supramarginal gyrus, inferior parietal lobule, insula, and inferior and superior frontal gyri (Schurz et al., 2014).

3. Methods

3.1. Data collection

The functional sector of the BrainMap database (Fox & Lancaster, 2002) was queried to identify fMRI experimental data related to the ToM function. A parallel standardized search was performed on BrainMap to retrieve fMRI experimental data related to other cognitive tasks stored in the database. The final literature search was obtained in December 2024, with no restrictions on publication year.

Eligible data (i.e., x-y-z foci of activation) were published in a peer-reviewed English-language article, statistically evaluating task-based brain activations related to groups of healthy human subjects and reported a whole-brain fMRI analysis with stereotactic results (i.e., Talairach or Montreal Neurological Institute standard spaces). The inclusion criteria were designed to mitigate biases inherent in region-of-interest analyses and to minimize spatial inaccuracies (Manuello et al., 2022).

3.2. Data organization

Two distinct datasets were created to estimate the selectivity of the cognitive function of interest: 1) "ToM dataset", composed of experimental data reporting brain activation in TOM, and 2) "non-ToM dataset", composed of experimental data reporting brain activations in all other tasks stored in the BrainMap database. Analyses were conducted in the Montreal Neurological Institute space. Thus, the spatial accuracy of the data was improved by converting foci reported in Talairach into Montreal Neurological Institute space using the icbm2tal algorithm (Lancaster et al., 2007).

3.3. Statistical analysis

The BACON approach (Costa et al., 2021) was applied to estimate the probability that brain activations are selectively associated with ToM.

First, two separate meta-analyses were conducted using the activation likelihood estimation (ALE) method (Eickhoff et al., 2012): one based on the "ToM dataset" and another using the "non-ToM dataset", which included all other tasks. The ALE algorithm, implemented in the GingerALE software (v.3.0.2) (Eickhoff et al., 2016), models the activation foci from each fMRI experiment as three-dimensional

Gaussian probability distributions centered on the reported activation foci. This process generates a modeled activation map for each experiment. The size of the Gaussian kernel varies to account for the original sample size of each group. The combination of all modeled activation maps yields voxelwise ALE scores across the whole brain, quantifying the degree of spatial overlap in reported activations.

Next, the BACON algorithm, as implemented in the MANGO software (v.4.1), was applied. By integrating Bayes Factor analysis (Kass & Raftery, 1995) with the unthresholded ALE-derived maps, BACON quantifies the posterior probability that activations at each brain voxel are selectively linked to the function of interest rather than to other cognitive tasks. This approach enabled a voxelwise whole-brain evaluation of two competing hypotheses: (1) that the activation was associated with ToM, or (2) that it was also linked to other experimental tasks of the BrainMap database. In the absence of prior probability estimates for these hypotheses, they were assumed to be equally likely, following previous validation studies (Cauda et al., 2020; Costa et al., 2021). Ultimately, BACON calculated posterior probabilities, representing *P (Theory-of-Mind | activation)*, to determine the selective association between observed brain activations and the ToM function. A detailed statistical explanation is provided in Costa et al. (2021).

Results were initially thresholded at *P (Theory-of-Mind | activation)* ≥ 0.90, corresponding to a posterior probability of selectivity of 90% or higher (Costa et al., 2021; Liloia et al., 2023). Given the exploratory nature of the analysis, results were also examined using more stringent thresholds of 0.95 (i.e., selectivity value of 95% or higher) and 0.99 (i.e., selectivity value of 99% or higher).

4. Results

The comprehensive search yielded a total of 8377 published fMRI experiments, including 111 different tasks. The distribution of the ToM dataset was 223 experiments, 1069 subjects, and 1526 activation foci. The non-ToM dataset was composed of 8154 experiments, 127112 subjects, and 66649 foci. For a complete list and description of fMRI experimental tasks stored in the BrainMap database, refer to https://brainmap.org/taxonomy/paradigms/ .

4.1. Selective activation profile of Theory-of-Mind

Taking into account a selectivity value of 90%, the BACON approach identified cortical and cerebellar activation areas related to ToM. Specifically, 12 clusters (*k* size > 150 mm^3) were found, encompassing the bilateral inferior, middle, and superior temporal gyri, superior frontal gyri, and cerebellar crus II. Additional selective activations were observed in the right anterior cingulate cortex and precuneus (Table 1, Figure 1).

Using a selectivity value of 95%, the BACON approach revealed one cortical area of activation in ToM showed a *k* size > 150 mm^3, encompassing the left middle temporal gyrus (Table 2 and Figure 1). In contrast, no ToM-related activations were found using a selectivity value of 99%.

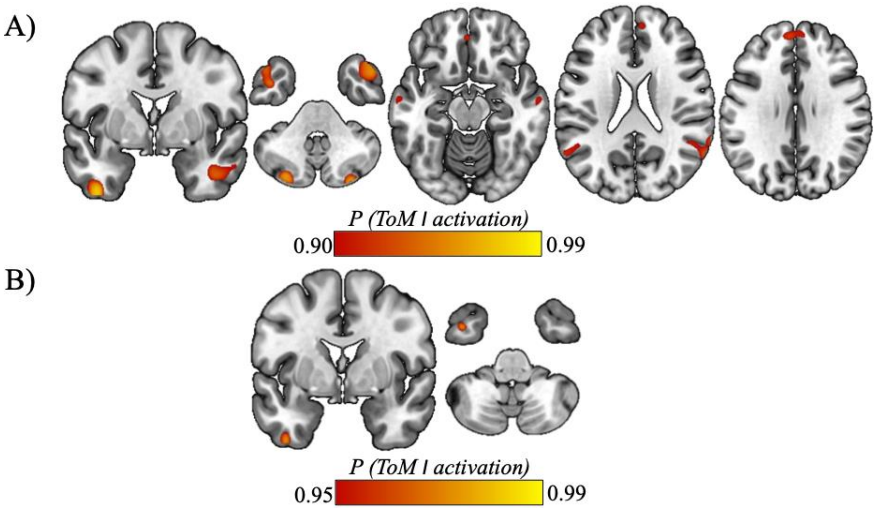
Table 1. Brain clusters of activation in Theory-of-Mind tasks derived from the Bayes fACtor mOdeliNg analysis thresholded at P (Theory-of-Mind | activation) ≥ 0.90.

Cluster ID	Brain Region (Brodmann area)	MNI x y z	Cluster Size mm^3	BACON Value Maximum	Minimum
1	Right middle temporal gyrus (BA 21)	48 10 -42	4540	0.95534	0.90001
2	Left middle temporal gyrus (BA 21)	-42 0 -46	1874	0.9711	0.90002
3	Left superior frontal gyrus (BA 8)	-6 52 32	1458	0.93052	0. 90001
4	Right cerebellar crus II	28 -82 -36	1253	0.95621	0. 90001
5	Right superior temporal gyrus (BA 39)	62 -60 22	964	0.93586	0.9
6	Left superior temporal gyrus (BA 22)	-60 -60 18	579	0.91656	0.9
7	Left cerebellar crus II	28 -88 -40	554	0.95984	0. 90005
8	Right anterior cingulate cortex (BA 32)	20 28 24	405	0.94271	0.9
9	Right precuneus (BA 7)	8 -56 32	338	0.91631	0.9
10	Right inferior temporal gyrus (BA 21)	62 -12 -18	313	0.91888	0. 90004
11	Left inferior temporal gyrus (BA 21)	-62 -10 -14	228	0.91193	0.9
12	Right superior frontal gyrus (BA 8)	20 42 18	154	0.92406	0. 90003

Table 2. Brain clusters of activation in Theory-of-Mind tasks derived from the Bayes fACtor mOdeliNg analysis thresholded at P (Theory-of-Mind | activation) ≥ 0.95.

Cluster ID	Brain Region (Brodmann area)	MNI x y z	Cluster Size mm^3	BACON Value Maximum	Minimum
1	Left middle temporal gyrus (BA 21)	-42 0 -46	267	0.9711	0.95001

Figure 1. Brain clusters of activation in Theory-of-Mind tasks derived from the Bayes fACtor mOdeliNg analysis thresholded at P (Theory-of-Mind | activation) ≥ 0.90 (A) and P (Theory-of-Mind | activation) ≥ 0.95 (B).



5. Discussion

Despite decades of neuroimaging research on brain–behavior relationships, a precise characterization of the possible selective function of brain regions, considering their involvement in multiple cognitive processes, remains elusive. Recent advancements in data aggregation methods have paved the way for data-driven, hypothesis-free approaches to mapping behavioral associations across brain regions.

In this study, we conducted an explorative investigation into the selective task-based activation profile of ToM using peer-reviewed fMRI data from the BrainMap database as the foundation for a whole-brain, voxelwise, and Bayesian analysis. Our meta-analytic approach identified multiple brain regions with a strong evidence of selective ToM activation compared with 110 other cognitive tasks. The functional localization of these areas highlights the involvement of specific cortical and cerebellar regions, including the bilateral inferior, middle, and superior temporal gyri, superior frontal gyri, and cerebellar crus II. In contrast, several areas traditionally associated with the "core ToM network" (i.e., posterior cingulate cortex, inferior frontal gyrus, insular cortex, left precuneus, inferior parietal lobule and angular gyrus) (Schurz et al., 2014) did not show posterior probability of selectivity at *P (Theory-of-Mind | activation)* ≥ 90%. This suggests that while these areas contribute to ToM processing, they are also engaged in a broader range of cognitive functions. Overall, this is not a surprising result given that previous fMRI findings support the view that several cortical and subcortical areas constitute crucial nodes of a multimodal network involved in a plethora of cognitive functions (Cauda et al., 2012). On the other hand, it is important to highlight that when increasing the posterior probability threshold for selectivity to a very high level of evidence (i.e., $P \geq 95\%$), only the left middle temporal gyrus remains selective. This finding suggests a central role for this multimodal area in ToM processing.

Several limitations should be acknowledged when interpreting these results. First, while there is no strong reason to assume systematic biases in the reporting of experiments, the BrainMap database used for dataset creation may not reflect the real-world distribution of fMRI tasks. Moreover, the design constraints of the original experiments limit the ability to explore potential differences across age- or sex-stratified populations. Finally, we cannot determine how many whole-brain fMRI studies may have overlooked the cerebellum, either partially or entirely, during scanning. As this study highlights, the cerebellum appears to play a significant role in cognitive functions and should be systematically included in fMRI acquisition and subsequent analyses.

Of course, this study represents only an initial step in the systematic exploration of *reverse inference* in cognitive neuroscience. The intent of this work is therefore programmatic. We argue that a more precise integration of *forward* and *reverse inference* could provide new insights, addressing key conceptual challenges, and fostering methodological advancements in fMRI research.

References

Cauda, F., Nani, A., Liloia, D., Manuello, J., Premi, E., Duca, S., Fox, P. T., & Costa, T. (2020). Finding specificity in structural brain alterations through Bayesian reverse inference. *Human Brain Mapping*, *41*(15), 4155-4172. https://doi.org/10.1002/hbm.25105

Cauda, F., Torta, D. M.-E., Sacco, K., Geda, E., D'Agata, F., Costa, T., Duca, S., Geminiani, G., & Amanzio, M. (2012). Shared «core» areas between the pain and other task-related networks. *PloS One*, *7*(8), e41929. https://doi.org/10.1371/journal.pone.0041929

Costa, T., Manuello, J., Ferraro, M., Liloia, D., Nani, A., Fox, P. T., Lancaster, J., & Cauda, F. (2021). BACON: A tool for reverse inference in brain activation and alteration. *Human Brain Mapping*, *42*(11), 3343-3351. https://doi.org/10.1002/hbm.25452

Eickhoff, S. B., Bzdok, D., Laird, A. R., Kurth, F., & Fox, P. T. (2012). Activation likelihood estimation meta-analysis revisited. *NeuroImage*, *59*(3), 2349-2361. https://doi.org/10.1016/j.neuroimage.2011.09.017

Eickhoff, S. B., Nichols, T. E., Laird, A. R., Hoffstaedter, F., Amunts, K., Fox, P. T., Bzdok, D., & Eickhoff, C. R. (2016). Behavior, Sensitivity, and power of activation likelihood estimation characterized by massive empirical simulation. *NeuroImage*, *137*, 70-85. https://doi.org/10.1016/j.neuroimage.2016.04.072

Fox, P. T., & Lancaster, J. L. (2002). Opinion: Mapping context and content: the BrainMap model. *Nature Reviews. Neuroscience*, *3*(4), 319-321. https://doi.org/10.1038/nrn789

Henson, R. (2006). Forward inference using functional neuroimaging: Dissociations versus associations. *Trends in Cognitive Sciences*, *10*(2), 64-69. https://doi.org/10.1016/j.tics.2005.12.005

Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, *90*(430), 773-795. https://doi.org/10.1080/01621459.1995.10476572

Lancaster, J. L., Tordesillas-Gutiérrez, D., Martinez, M., Salinas, F., Evans, A., Zilles, K., Mazziotta, J. C., & Fox, P. T. (2007). Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. *Human Brain Mapping*, *28*(11), 1194-1205. https://doi.org/10.1002/hbm.20345

Liloia, D., Cauda, F., Uddin, L. Q., Manuello, J., Mancuso, L., Keller, R., Nani, A., & Costa, T. (2023). Revealing the Selectivity of Neuroanatomical Alteration in Autism Spectrum Disorder via Reverse Inference. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *8*(11), 1075-1083. https://doi.org/10.1016/j.bpsc.2022.01.007

Manuello, J., Costa, T., Cauda, F., & Liloia, D. (2022). Six actions to improve detection of critical features for neuroimaging coordinate-based meta-analysis preparation. *Neuroscience & Biobehavioral Reviews*, *137*, 104659. https://doi.org/10.1016/j.neubiorev.2022.104659

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, *10*(2), 59-63. https://doi.org/10.1016/j.tics.2005.12.004

Poldrack, R. A. (2008). The role of fMRI in Cognitive Neuroscience: Where do we stand? *Current Opinion in Neurobiology*, *18*(2), 223-227. https://doi.org/10.1016/j.conb.2008.07.006

Poldrack, R. A., Kittur, A., Kalar, D., Miller, E., Seppa, C., Gil, Y., Parker, D., Sabb, F., & Bilder, R. (2011). The Cognitive Atlas: Toward a Knowledge Foundation for Cognitive Neuroscience. *Frontiers in Neuroinformatics*, *5*. https://www.frontiersin.org/articles/10.3389/fninf.2011.00017

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515-526. https://doi.org/10.1017/S0140525X00076512

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience and Biobehavioral Reviews*, *42*, 9-34. https://doi.org/10.1016/j.neubiorev.2014.01.009

Westbrook, C. (2021). *Handbook of MRI Technique*. John Wiley & Sons.