

RESEARCH ARTICLE

ROI and phobias: The effect of ROI approach on an ALE meta-analysis of specific phobias

Claudio Gentili¹  | Simone Messerotti Benvenuti¹  | Giada Lettieri²  |
Cristiano Costa¹  | Luca Cecchetti² 

¹Department of General Psychology,
University of Padova, Padova, Italy

²IMT School for Advanced Studies Lucca,
Lucca, Italy

Correspondence

Claudio Gentili, MD, PhD, Department of
General Psychology, University of Padova, Via
Venezia 8, 35131, Padova, Italy.
Email: c.gentili@unipd.it

Abstract

About 90% of fMRI findings on specific phobias (SP) include analysis of region of interest (ROI). This approach characterized by higher sensitivity may produce inflated results, particularly when findings are aggregated in meta-analytic maps. Here, we conducted a systematic review and activation likelihood estimation (ALE) meta-analysis on SP, testing the impact of the inclusion of ROI-based studies. ALE meta-analyses were carried out either including ROI-based results or focusing on whole-brain voxelwise studies exclusively. To assess the risk of bias in the neuroimaging field, we modified the Newcastle–Ottawa Scale (NOS) and measured the reliability of fMRI findings. Of the 31 selected investigations (564 patients and 485 controls) one-third did not motivate ROI selection: five studies did not report an explicit rationale, whereas four did not cite any specific reference in this regard. Analyses including ROI-based studies revealed differences between phobics and healthy subjects in several regions of the limbic circuit. However, when focusing on whole-brain analysis, only the anterior midcingulate cortex differentiated SP from controls. Notably, 13 studies were labeled with low risk of bias according to the adapted NOS. The inclusion of ROI-based results artificially inflates group differences in fMRI meta-analyses. Moreover, a priori, well-motivated selection of ROIs is desirable to improve quality and reproducibility in SP neuroimaging studies. Lastly, the use of modified NOS may represent a valuable way to assess and evaluate biases in fMRI studies: “low risk” of bias was reported for less than half of the included studies, indicating the need for better practices in fMRI.

KEYWORDS

ALE meta-analysis, fMRI, phobias, region of interest

1 | INTRODUCTION

Specific phobias (SP) are defined as fear or anxiety about a particular object or situation that generates an immediate reaction in the individual, who will therefore constantly (try to) avoid the object or the situation to reduce the sense of intense unpleasantness (APA, 2013).

Phobias have been extensively studied using in vivo neuroimaging techniques, since simple experimental designs can efficiently tap into the relatively low clinical complexity of this condition. Specifically, typical paradigms consist in measuring brain activity of a subject exposed either to phobic stimuli, stimuli considered generally aversive or, although more rarely, nonaversive stimuli with positive or negative emotional connotation. For instance, earliest studies simply evaluated

whether brain response to phobic stimuli was different in phobic patients as compared with healthy controls (HC; e.g., Dilger et al., 2003). Further investigations explored different aspects of phobias, including the evaluation of different subtypes (e.g., blood, injection, or dental phobia; Caseras, Giampietro, et al., 2010; Halsband & Wolf, 2015) and the use of different presentation modalities, such as written words (e.g., Straube, Mentzel, Glauer, & Miltner, 2004). Other lines of research examined whether, following treatment, brain responses of phobic patients matched the brain activity of healthy participants (e.g., Lipka, Hoffmann, Miltner, & Straube, 2014; Schienle, Schafer, Hermann, Rohrmann, & Vaitl, 2007; Straube, Glauer, Dilger, Mentzel, & Miltner, 2006), or whether certain subgroups of phobic patients responded in different ways to phobic stimuli (e.g., small animals

vs. spider phobias; Caseras, Mataix-Cols, et al., 2010; Lueken et al., 2011, 2014). Other studies also focused on classical conditioning in phobic patients as one of the most important mechanisms involved in the genesis of phobias (e.g., Goossens, Sunaert, Peeters, Griez, & Schruers, 2007b; Schweckendiek et al., 2011). Moreover, by recording brain responses to aversive nonphobic and to emotionally connoted stimuli, several studies have assessed whether individuals with phobia process nonphobic stimuli differently as compared with healthy subjects (e.g., emotional faces; Britton, Gold, Deckersbach, & Rauch, 2009; Killgore et al., 2014; Wright, Martis, McMullin, Shin, & Rauch, 2003).

A qualitative evaluation of this large body of literature reveals the consistent alteration of a widespread brain network in phobic patients, encompassing the amygdala, the insular cortex, the cingulate cortex, part of the occipital visual areas and the prefrontal cortex. Changes in this network typically occur across different types of phobias and stimuli (e.g., phobic, adverse nonphobic stimuli, and faces), and are modulated by treatment (for a narrative synthesis of these findings see Del Casale et al., 2012). However, to the best of our knowledge, no meta-analysis has been conducted in this regard.

Studies examining brain activity in individuals with phobias have frequently used region of interest (ROI) analysis, a popular method to spatially limit the search for significant differences in neuroimaging studies, which, however, is justified only when a strong a priori hypothesis is present (Poldrack, 2007). A reason for the adoption of the ROI approach in phobias research is its ability to maximize the identification of differences when focusing on smaller brain regions (e.g., the amygdala). Indeed, due to their intrinsic morphological properties, smaller regions may not survive the typical cluster-based correction for multiple comparisons since their volume could be smaller than the minimum cluster size computed on the whole-brain.

However, the use of ROI approach to merely overcome this limitation is not fully justified, given the recent advances in methods for the assessment of statistical significance. In particular, although voxel-based correction methods (e.g., false discovery rate, Bonferroni) are well-known to inflate Type II error (Lieberman & Cunningham, 2009), other "hybrid" approaches, such as the threshold-free cluster enhancement (TFCE) algorithm, are less conservative as compared with "pure" voxel-based methods, but retain the ability to capture activations in smaller brain regions (Smith & Nichols, 2009). In addition, albeit the use of a priori hypotheses to limit the search for significance is not a caveat per se, the activity of some brain areas not included in the preselected volume of interest could be actually related to phobias, but neglected because of the ROI approach. Furthermore, to delineate ROIs borders, a possible way is to take advantage of brain anatomy (Poldrack, 2007): researchers can either manually trace ROI using landmarks based on each subject peculiar anatomy, or use standard space atlases. However, both these methods are based on the strong assumption that a significant overlap between brain structure and function exists and that this correspondence is approximately constant across different cortical and subcortical areas (Huettel, Song, & McCarthy, 2004).

All in all, one may argue that the use of ROIs is particularly useful to increase the chances to detect the effect of interest, especially for studies based on smaller samples. As the p -value depends on both the magnitude of the effect and the sample size, if one assumes the effect

size to be constant it is clear that statistical significance after correction for multiple comparisons (i.e., dividing the p -value by the number of tests) requires a large number of subjects to be acquired. This is what happens in mass-univariate voxelwise testing, whereas in ROI analysis the comparison between two conditions or groups is performed on average ROI activity and, thus, only a single test is carried out. The same also holds when ROIs are used to simply restrict the search for significance, as in the case of small-volume correction. Here, the number of tests is reduced as, rather than considering all voxels, the analysis is limited to few hundreds or thousands (i.e., those included in the ROI), resulting in a more lenient corrected threshold. However, all this comes at the cost of increasing Type I error, especially when more than one region is considered and no (or questionable) a priori criteria guide the ROI selection. Importantly, no specific methods have been developed to assess study quality and risk of bias (RoB) in this regard, even though these issues are crucial when it comes to aggregate findings at meta-analytic level.

In addition, a very recent article (Müller et al., 2018) questioned the inclusion of ROI-based studies in meta-analyses as this approach could lead to inflated results, thus raising also the question of how findings obtained from these experiments should be aggregated. Therefore, we conducted a systematic review and activation likelihood estimation (ALE) meta-analysis on SP studies, either including or excluding studies based on ROI analyses, along with the voxelwise whole-brain approach. Our specific aims were the following:

1. To summarize studies examining brain activity elicited by visual phobia-related stimuli in individuals with SP.
2. To systematically evaluate the use and the selection of ROIs in phobia studies.
3. To assess the impact of the ROI approach on meta-analytic activation patterns.
4. To provide a scale assessing the RoB of functional magnetic resonance imaging (fMRI) studies, which may also be used in future systematic reviews and neuroimaging meta-analysis.

2 | METHODS

2.1 | Study selection

We considered fMRI studies that assessed brain responses (in terms of signal change) of patients with SP and controls elicited by visual stimuli. To do so, we performed a search on the Pubmed bibliographic database using the keyword "fMRI" together with the terms: "specific phobia," "simple phobia," "blood phobia," "injection phobia," "animal phobia," "arachnophobia," "spider phobia," and "dental phobia." This search was performed on 7th January 2017 and the research string is reported in the Supporting Information material.

Two of the authors (CG and CC) thoroughly screened and independently selected all the studies in which: (1) a visual stimulus was presented during the fMRI scan; (2) a direct comparison of brain activation between individuals with phobias and controls was included (i.e., HC > SP and/or SP > HC); (3) the coordinates of this comparison

were reported. Data selection and data analysis were registered on the platform PROSPERO (<https://www.crd.york.ac.uk/PROSPERO/>; registration number: CRD42018084940).

2.2 | Studies quality

We evaluated the quality and RoB of the selected studies. Since, to the best of our knowledge, there is no standardized tool for assessing these important issues in fMRI meta-analyses, a modified version of the Newcastle–Ottawa scale was created (NOS; Stang, 2010). Specifically, three of the NOS domains were adapted to fMRI data for the purpose of the present study: (1) the quality of sample selection was assessed based on (i) case definition (i.e., the method used to select individuals with phobias), (ii) representativeness (i.e., the method used to make the sample representative of the population), (iii) selection of the control group (i.e., the method used to select healthy individuals assigned to the control group), and (iv) definition of controls (i.e., which were the inclusion/exclusion criteria for controls selection); (2) the quality of reproducibility was evaluated considering whether the samples were comparable in terms of age and in terms of other variables considered in the study; (3) the quality of exposure was assessed based on (i) whether the two groups underwent the same experimental procedure, (ii) whether the article described drop-outs, and (iii) whether a behavioral effect was measurable (e.g., subjective higher arousal rates for phobic images in patients as compared with controls). Lastly, a fourth domain for (4) quality in data analysis was introduced. In particular, we evaluated (i) whether the studies used a sufficient cluster-forming threshold (i.e., voxel-based uncorrected p -value) to compensate for false positive underestimation, as recently described (Eklund, Nichols, & Knutsson, 2016), and (ii) whether the authors used and described a valid correction method to avoid/reduce false positive results. This latter domain may be controversial, since the inflated false positive issue, as well as possible solutions, have not been debated until recently (Cox, Reynolds, & Taylor, 2016; Eklund et al., 2016). Notwithstanding, these information should be taken into account when estimating the reliability of findings coming from fMRI studies (Cremers, Wager, & Yarkoni, 2017). A comprehensive description of these procedures is reported in the Supporting Information Methods section.

Lastly, two of the authors (CG, CC) assessed the RoB for all the selected studies. Inter-rater agreement κ statistic was computed, and divergence of opinion was resolved by discussion between assessors. The modified version of NOS provided a score ranging from 0 to 11. We considered scores between 0 and 3 as indicative of high risk, scores between 4 and 7 as intermediate and scores between 8 and 11 as low. A detailed description of this new tool can be found in Supporting Information Materials (Supporting Information Table S1).

2.3 | Study overlap analysis

For studies authored by the same research group, we personally contacted the corresponding authors asking whether there was any overlap in the sample and, if so, the degree of such overlap. Also, to mitigate the effects of dependency in sample selection, we ran two complementary analyses and tested the robustness of the results. In

the first analysis we pooled in a single experiment all the results coming from studies with overlap among subjects and sample size was reduced accordingly (pooled-studies analysis). Specifically, if experiment A included $N = 20$ subjects and experiment B included $N = 10$ but the two shares 50% of the subjects, in this first analysis only one experiment is considered and sample size is set to $N = 25$. In the second analysis (leave-studies out analysis), we simply excluded all the smaller studies with overlap among subjects, maintaining only the study with the largest number of participants. For instance, if 20 subjects took part in experiment A but 5 of them participated also in experiment B (where $N = 10$), then experiment B is excluded from the analysis and only coordinates coming from experiment A are considered ($N = 20$) (see Supporting Information Materials for further details).

2.4 | ALE meta-analysis

Here, we used GingerALE 2.3.6 (<http://www.brainmap.org/ale/>; Eickhoff et al., 2009) to perform six different ALE meta-analyses (three conditions, i.e., voxelwise whole-brain only, ROI only, and combination of ROI and whole-brain, for each of the two contrasts of interest, i.e., $SP > HC$, $HC > SP$). For the first two maps we included coordinates from studies based on the voxelwise whole-brain analysis method only and relative to $SP > HC$ and $HC > SP$ separately. For the next two maps, we selected coordinates from studies based on ROI analysis only and the $SP > HC$ as well as $HC > SP$ contrasts, whereas for the last two we employed results obtained from all the studies regardless of the analysis pipeline and for both contrasts of interest. Importantly, when two analysis methods were employed, we pooled in a single experiment coordinates from ROI and whole-brain studies to adjust for multiple contrasts/comparisons, as in Turkeltaub et al. (2012). For each study, we included all the contrasts (e.g., spider vs. neutral/spider vs. snake/snake vs. neutral) and experiments (e.g., passive viewing/active detection) in which a visual stimulus is used and the contrasts $SP > HC$ and/or $HC > SP$ are reported. We did not include results in which participants were in altered state of consciousness (e.g., hypnosis) or in which the same dataset is analyzed in different ways. In order to control for the inclusion of multiple comparisons derived from the same dataset, the correction proposed by Turkeltaub et al. (2012) and implemented in GingerALE was used. Therefore, the number of experiments entered in our meta-analysis is equal to the number of the studies/articles considered. Supporting Information Table S2 describes in detail which contrasts have been used for the meta-analyses.

Coordinates in the MNI 152 standard space were converted into the Talairach space using the GingerALE foci converter tool.

The statistical significance was assessed and corrected for multiple comparisons employing a cluster-based method: $p < 0.001$ cluster forming threshold, $p < 0.05$ cluster corrected FWE and $N = 2,000$ permutations.

For each of the six meta-analyses we considered the regions reported for the contrast $SP > HC$ and $HC > SP$ separately. Moreover, the overlap between $SP > HC$ and $HC > SP$ maps was computed employing AFNI (Cox, 1996): for each condition of interest (i.e., whole-brain, combined and ROI) and each contrast (i.e., $HC > SP$

and SP > HC) we first identified significant clusters using GingerALE. Afterward, we binarized and assigned a dummy value (i.e., ID) to each of these maps depending on the contrast: 1 for HC > SP and 2 for SP > HC. Lastly, we added the HC > SP dummy contrast map to the SP > HC one within each condition of interest, generating three final images, where voxels having intensity 1 are significant for the HC > SP contrast exclusively, those having intensity 2 are significant for the SP > HC and the ones with intensity 3 are significant for both contrasts.

2.5 | Omnibus effect ALE meta-analysis

After careful evaluation of coordinates reported in the selected manuscripts, we noticed that few studies included results for the HC > SP contrast (i.e., only 10 experiments). Therefore, we contacted the corresponding authors of all the studies where this contrast was not mentioned and asked whether the analysis produced significant results for HC > SP that were, however, not reported in the manuscript (see section on "Characteristics of Included Studies" for further details).

In addition, since the number of experiments was not sufficient to guarantee the validity of the HC > SP results (Müller et al., 2018), we computed an omnibus ALE meta-analysis test. In this regard, we pooled together results of SP > HC and HC > SP as in an omnibus test (SP ≠ HC), answering the question whether there were areas of convergence of differences in brain activation between groups, regardless of the directionality. We performed this analysis for each of the three conditions (i.e., voxelwise whole brain only, ROI only and combination of ROI and whole-brain). Since this analysis was decided after extracting the coordinates from all the selected manuscripts, it is not described in the trial registration.

3 | RESULTS

3.1 | Selection and inclusion of studies

Our search string produced 190 entries. Starting from this pool of results, we discarded studies on other mental or somatic conditions, researches with no control group, those in which a visual task was not present, functional connectivity analysis only and reviews. Of the 42 remaining studies, 11 were further discarded because of one of the following reasons: (1) no direct comparison between individuals with phobias and controls; (2) no coordinates for contrasts or ROI reported; and (3) re-analyses of previous articles (see Figure 1).

3.2 | Characteristics of included studies

The 31 studies included 564 patients with SP and 485 HC (Brinkmann, Poller, Herrmann, Miltner, & Straube, 2017; Britton et al., 2009; Caseras, Giampietro, et al., 2010; Caseras, Mataix-Cols, et al., 2010; Dilger et al., 2003; Goossens, Schruers, Peeters, Griez, & Snaert, 2007a; Goossens et al., 2007b; Halsband & Wolf, 2015; Herrmann et al., 2007; Hilbert, Evens, Maslowski, Wittchen, & Lueken, 2014; Killgore et al., 2014; Lipka et al., 2014; Lipka, Miltner, & Straube, 2011; Lueken et al., 2011, 2014; Michalowski et al., 2017; Munsterkotter et al., 2015; Scharmuller et al., 2014; Schienle et al.,

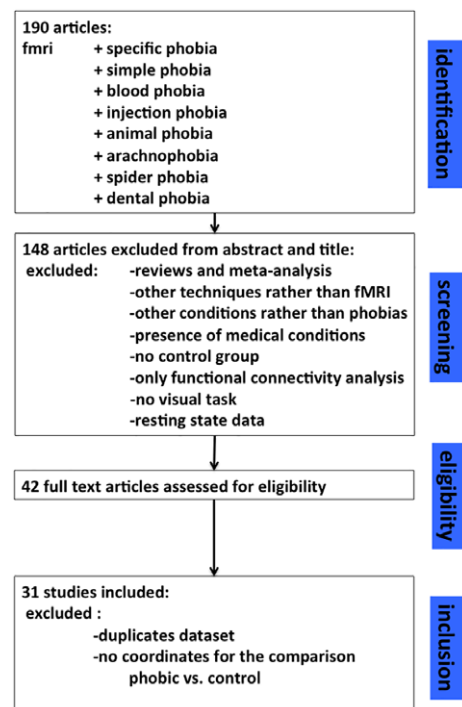


FIGURE 1 Prisma graph illustrating the selection process of the present meta-analysis [Color figure can be viewed at wileyonlinelibrary.com]

2007; Schienle, Schafer, Walter, Stark, & Vaitl, 2005; Schienle, Scharmuller, Leutgeb, Schafer, & Stark, 2013; Schweckendiek et al., 2011; Straube, Glauer, et al., 2006; Straube, Lipka, Sauer, Mothes-Lasch, & Miltner, 2011; Straube et al., 2004; Straube, Mentzel, & Miltner, 2006, 2007; Wendt, Lotze, Weike, Hosten, & Hamm, 2008; Wiemer et al., 2015; Wright et al., 2003; Zilverstand, Sorger, Kaemingk, & Goebel, 2017). In six studies, two different subtypes of phobias were considered, with the most studied (24 out of 31) subtype being spider phobia; dental phobia was evaluated in six studies, whereas small animal and blood/injection subtypes in four and three studies, respectively (Table 1). Twenty-one studies performed both ROI and voxel-wise whole-brain analysis, whereas seven used the ROI approach only and the remaining three used the whole-brain approach exclusively (Supporting Information Figure S1).

As far as ROIs are concerned, five studies did not report an explicit rationale for the selection of the ROIs. For the other 23 studies, ROIs were selected in accordance with previous findings on either activations in individuals with phobias or on the processing of threatening and fearful images in healthy volunteers. However, only 19 of these latter studies reported references to previous articles to justify their choice.

About 10 studies reported results for the HC > SP comparison, further 10 no significant results, whereas 11 studies did not provide any information in this regard. To further explore whether significant results for the comparison HC > SP were found but not reported, we personally contacted the corresponding authors of the selected publications. The seven responses we received confirmed that no other significant results for the HC > SP comparisons were found (Supporting Information Table S3).

TABLE 1 Characteristics of the studies included in the meta-analysis

Experiment	Phobics				Healthy controls			Task and stimuli	Whole brain or ROI
	N.	Type	M/F	Age	N.	M/F	Age		
Brinkmann et al., 2017	16	BII	0/16	24.1 ± 3.82	16	0/16	23.7 ± 4.44	Anticipation and exposure of phobia-specific and neutral pictures	ROI
Britton et al., 2009	12	SAP	5/7	25.2 ± 4.5	12	4/8	26.7 ± 5.5	Emotional Stroop task with phobic and neutral words	Both
Caseras, Giampietro, et al., 2010	14	SP	2/12	21.50 ± 2.73	17	2/15	21.76 ± 2.88	Exposure to phobic and nonphobic stimuli	Both
	15	BII	2/13	22.40 ± 2.32					
Caseras, Mataix-Cols, et al., 2010	14	SP	3/11	22.71 ± 2.89	14	3/11	23 ± 2.66	Exposure to phobic and nonphobic stimuli	Whole brain
	12	BII	3/9	24.58 ± 4.31					
Dilger et al., 2003	9	SP	0/9	25 ± 2.3	9	0/9	21.3 ± 0.6	Exposure to phobic, aversive (snakes) and nonphobic (mushrooms) stimuli	Both
Goossens et al., 2007a	15	SP	2/13	24 ± 2	14	2/12	23 ± 1	Exposure to phobic, aversive (snakes) and nonphobic (mushrooms) stimuli	Both
Goossens et al., 2007b	16	SP	0/16	24 ± 3.02	14	2/12	24 ± 3.02	Exposure to phobic, aversive (snakes) and nonphobic (mushrooms) stimuli	Both
Halsband & Wolf, 2015	12	DP	4/8	34.9	12	6/6	33.15	Exposure to audio-visual phobic stimuli	ROI
Hermann et al., 2007	9	BII	0/9	22.9 ± 4.7	10	0/10	27.6 ± 10.6	Exposure to phobic, disgusting, fearful and neutral stimuli	Both
Hilbert et al., 2014	13	DP	4/9	24.92 ± 2.25	13	4/9	23.23 ± 3.19	Exposure to audio-visual phobic stimuli	Whole brain
Killgore et al., 2014	15	SAP	4/11	35.6 ± 8.7	22	8/14	30.7 ± 9.2	Exposure to masked emotionally expressive and neutral faces	ROI
Lipka et al., 2011	18	SP	0/18	25.56 ± 5.26	18	0/18	24.72 ± 5	Exposure to masked phobic and neutral stimuli	ROI
Lipka et al., 2014	14	SP	0/14	23.57 ± 4.42	14	0/14	25.00 ± 3.70	Exposure to Subliminal and supraliminal phobic and neutral stimuli	Both
	14	SP	0/14	23.88 ± 3.65					
Lueken et al., 2011	12	SP	3/9	25.6 ± 7.54	17	5/12	23.7 ± 4.44	Exposure to phobic and neutral video stimuli	Both
	12	DP	3/9	25.1 ± 7.03					
Lueken et al., 2014	13	SP	2/11	21.85 ± 1.95	13	3/10	21.46 ± 1.85	Anticipation and perception phase of phobic and neutral stimuli	Both
	13	DP	3/10	23.00 ± 3.37					
Michalowski et al., 2017	12 12	SP DP	2/10 1/11	22.8 ± 3.3 (data for single groups of phobia were not presented)	13	6/7	22.8 ± 3.3	Exposure to phobic and neutral pictures	Both
Munsterkotter et al., 2015	25	SP	2/23	24.4 ± 5.42	26	5/21	25.5 ± 5.13	Exposure to phobic and neutral stimuli under conditions of either predicted (phasic) or unpredicted (sustained) fear	ROI
Scharmuller et al., 2014	20	DP	0/20	28.7 ± 9.2	20	0/20	25.4 ± 8.0	Exposure to phobic and neutral images	Both
Schienenle et al., 2005	10	SP	0/10	22.5 ± 2.2	13	0/13	23.9 ± 6.8	Exposure to phobic, disgusting, fearful and neutral stimuli	Both
Schienenle et al., 2007	26	SP	0/26	27.2 ± 9.2 24.3 ± 2.0	25	0/25	24.6 ± 6.3	Exposure to phobic, disgusting, fearful and neutral stimuli	Both
Schienenle et al., 2013	45	SP	20/25	31.5 ± 10.5	41	18/23	30.9 ± 8.5		Both

(Continues)

TABLE 1 (Continued)

Experiment	Phobics				Healthy controls			Task and stimuli	Whole brain or ROI
	N.	Type	M/F	Age	N.	M/F	Age		
				29.5 ± 10.7			28.4 ± 9.4	Exposure to phobic, disgusting, fearful and neutral stimuli	
Schweckendiek et al., 2011	15	SP	2/13	23.53 ± 3.27	14	2/12	23.64 ± 3.43	Exposure to phobic and neutral stimuli	Both
Straube et al., 2004	11	SP	0/11	20.8	11	0/11	22.4	Exposure to phobia-relevant and neutral words	ROI
Straube, Mentzel, et al., 2006	28	SP	0/28	22.07 ± 1.98	14	0/14	22.07 ± 1.98	Exposure to phobic and neutral video stimuli	Both
Straube, Glauer, et al., 2006	11	SP	0/11	20.9 ± 2.3	12	0/12	21.3 ± 0.6	Exposure to phobic and neutral images	ROI
Straube et al., 2007	16	SP	0/16	21.8 ± 0.6	15	0/15	22.7 ± 0.9	Anticipation and presentation of phobic and neutral stimuli	Both
Straube et al., 2011	17	SP	0/17	25.2 ± 4.9	16	0/16	26.6 ± 9.2	Exposure to phobic and neutral images while performing a distraction task	Both
Wendt et al., 2008	13	SP	0/13	23.2	13	0/13	21.1	Exposure to phobic and neutral stimuli	Both
Wiemer et al., 2015	18	SP	0/18	21.4 ± 4.2	18	0/18	22.2 ± 2.2	Exposure to phobic and neutral stimuli followed randomly by a painful electrical shock	Both
Wright et al., 2003	10	SAP	4/6	29.8 ± 6.8	10	4/6	29.8 ± 6.8	Exposure to emotionally expressive and neutral faces	Both
Zilverstand et al., 2017	7	SP	0/7	21.7 ± 3.9	7	0/7	20.9 ± 2.2	Exposure to phobic stimuli in different contexts and at different zoom	Whole brain

SAP = small animal phobia; SP = spider phobia; BII = blood, injection and injury phobia; DP = dental phobia.

3.3 | Study quality

Study quality is summarized in Supporting Information Figure S2 and Supporting Information Table S1. Twenty-nine studies provided an adequate case definition, whereas no study fulfilled the criteria for population representativeness of the sample. About 22 studies provided an adequate selection of controls and 25 a satisfactory definition of criteria for the identification of control participants. Considering the comparability, age was matched across groups in 28 studies, whereas other variables were controlled in 22 studies. In 26 studies the groups underwent the very same experimental procedure. In the remaining 5 studies out of 31, differences in the experimental procedure were related to compensations for participating in the study, where for instance the phobic group benefited from a free therapy session, whereas the control group was rewarded with monetary incentives or no remuneration. In only 13 studies a documentation of dropouts was reported, and 23 measured a behavioral effect. Regarding the statistical quality, the adoption of the recently proposed cluster-forming threshold to control for false positive results in neuroimaging studies ($p < 0.001$) was found in only five studies (four studies provided the corrected p -value only—Supporting Information Table S3). Twenty-four studies clearly reported the methodology for multiple comparisons correction. According to our criteria,

13 studies were considered as having low RoB (maximum score 10), 18 were considered as having intermediate risk and no one with high risk (Supporting Information Table S3). Notably, no study scored the maximum of 11 points on our scale for the assessment of quality.

3.4 | ALE results

3.4.1 | Combined ROI and voxelwise whole-brain analysis

For the combined analysis, all the 31 studies were considered: 30 studies included coordinates for SP > HC contrast, whereas 10 for the HC > SP contrast (Table 1; Supporting Information Table S2). The minimum cluster size was 448 mm³ for the HC > SP contrast and 704 mm³ for the SP > HC contrast.

ALE meta-analysis results highlighted convergence of activation across studies for individuals with phobias in the amygdala, the insula, the anterior part of the midcingulate cortex extending to the medial portion of the superior frontal gyrus, the thalamus, the inferior frontal gyrus, and the basal ganglia, bilaterally (Figure 2a; Supporting Information Figure S3A; Table 2). The number of included experiments for the HC > SP comparison was smaller than the limit suggested by Eickhoff (2016); therefore, the results for this contrast should be considered exploratory. Convergence of activation for healthy controls was found

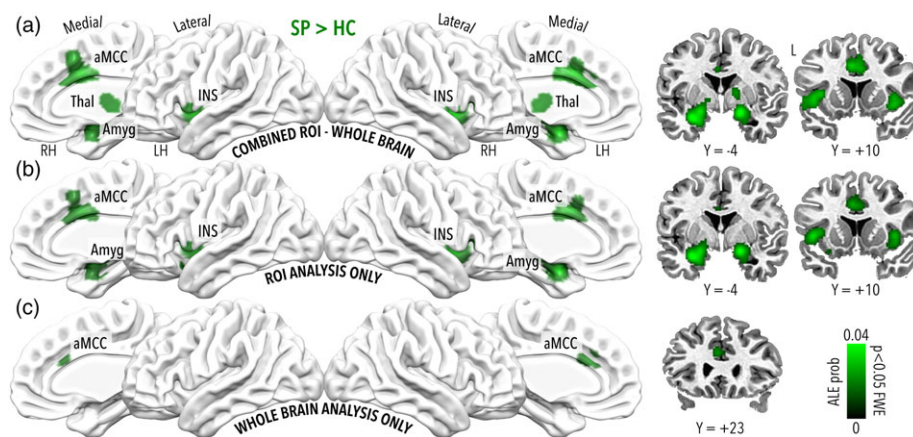


FIGURE 2 Significant results for the SP > HC contrast of interest (ALE $p < 0.05$ corrected). Panel a depicts clusters significant for the combined whole brain and ROI analysis, panel b for ROI analysis only and panel c for the voxelwise whole brain analysis only. Considering studies based on combined (panel a) and ROI analyses (panel b), convergence of activation was found in the amygdala, the insula extending to the medial portion of the superior frontal gyrus, the anterior part of the midcingulate cortex extending to the inferior frontal gyrus, the thalamus, and the basal ganglia. On the contrary, when focusing on studies based on voxelwise whole brain analysis (panel c) the only significant cluster was found in the anterior portion of the midcingulate cortex [Color figure can be viewed at wileyonlinelibrary.com]

in the ventral portion of the cingulate cortex extending to the orbitofrontal cortex, as well as in a posterior region of the left amygdala with a partial overlap with the results obtained from the SP > HC contrast (Supporting Information Figures S3A and S4A; Table 2).

3.4.2 | ROI analysis

Of the 28 studies included in the meta-analysis and based on the ROI approach, 26 reported coordinates for the SP > HC contrast and seven for the HC > SP contrast (Table 1; Supporting Information Table S2). The minimum cluster size was 552 mm^3 for the HC > SP contrast and 672 mm^3 for the SP > HC contrast.

Overall, ALE maps obtained for the two meta-analytic contrasts were similar to those derived pooling together all the 31 studies. Specifically, for the SP > HC contrast, activations pooled across studies converged in the bilateral amygdala, the inferior frontal gyrus and the anterior portion of midcingulate cortex, as well as in the right claustrum, the right medial frontal gyrus, the left insula, and parahippocampal gyrus (Figure 2b; Supporting Information Figure S3B; Table 2). The number of experiments for the HC > SP contrast is not adequate to ensure the robustness of the results, therefore findings for this contrast should be considered exploratory. Convergence of activation for healthy controls was found in right ventral cingulate cortex extending to the orbitofrontal cortex and in the left amygdala (Supporting Information Figures S3B and S4B; Table 2).

ROIs included for analysis in each of the selected studies and those found to be significantly different between groups are summarized in Table 3. Of note, bilateral amygdala was considered in 26 studies and significantly modulated by SP in 16 of those, whereas 10 studies reported only unilateral activations. Anterior cingulate regions were considered in 20 studies, with significant results reported 11 times. About 24 studies focused on the left insula, resulting significant for the contrasts of interest 14 times. The inferior frontal gyrus was considered in only two studies demonstrating significant results in both cases. The parahippocampus was included as ROI in two studies with one significant result. Right basal ganglia

were considered in two studies, both reporting significant results. Lastly, other ROIs were considered in the selected studies such as orbitofrontal cortex 12 times and fusiform gyrus 7 times (Table 3).

3.4.3 | Voxelwise whole-brain analysis

For the whole-brain analysis 19 studies were considered: 17 reported coordinates for the SP > HC contrast and 5 for the HC > SP contrast (Table 1; Supporting Information Table S2). Therefore, for the HC > SP comparison the number of included experiments fell below the limit (17–20 studies) suggested by Eickhoff (2016). The minimum cluster size was 640 mm^3 for the HC > SP contrast and 592 mm^3 for the SP > HC contrast.

The only significant cluster of convergence in activation was found in the anterior cingulate cortex for the SP > HC comparison (Figure 2c; Supporting Information Figure S3C; Table 2).

3.4.4 | Omnibus analysis

For each of the three conditions, results of the omnibus meta-analysis test (SP \neq HC) revealed an almost perfect overlap with the pattern of significant clusters obtained from the SP > HC contrast (Figure 3; Table 2). The only relevant difference was found for the ROI only condition, where a cluster extending to the left thalamus was present in the omnibus test (Figure 3b) but not in the SP > HC contrast (Figure 2b). Of note, the only significant result found for the HC > SP contrast and located in the ventral portion of the anterior cingulate cortex was not present in the omnibus test.

3.4.5 | Study overlap analysis

To assess the degree of overlap in sample across studies, we carefully checked the manuscripts and personally contacted the corresponding authors of researches included in our meta-analysis. We found that Goossens et al. employed the same control group (but different phobic individuals) across two different investigations (Goossens et al., 2007a; Goossens et al., 2007b). Also, the following authors reported to have included a subsample of subjects in two or more articles. The

TABLE 2 Talairach coordinates for the three ALE meta-analysis (whole-brain analysis alone, combined whole-brain analysis and ROI, and ROI alone) for SP > HC, HC > SP and SP ≠ HC

Whole brain	Hemisphere	Region	BA	Center of mass			Peak			ALE value at peak	Volume (mm ³)
				X	Y	Z	X	Y	Z		
SP > HC											
	L	Anterior cingulate	24	-2.6	23.9	25.2	-2	24	24	0.022	760
HC≠SP											
	L	Anterior cingulate	24	-2.7	24	25.2	-2	24	24	0.022	696
Whole brain +ROI											
Whole brain +ROI	Hemisphere	Region	BA	Center of mass			Peak			ALE value at peak	Volume (mm ³)
				X	Y	Z	X	Y	Z		
SP > HC											
	R	Amygdala		24.8	.4	-4.5	18	-4	-12	0.043	8,992
	R	Insula	13				38	6	0	0.034	
	R	Inferior frontal gyrus	47				42	20	2	0.026	
	L	Medial dorsal nucleus of thalamus					-2	-18	4	0.024	
	R	Ventral anterior nucleus of thalamus					14	-4	8	0.023	
	L	Medial dorsal nucleus of thalamus					-4	-14	8	0.023	
	R	Medial dorsal nucleus of thalamus					4	-14	10	0.018	
	L	Amygdala		-22.1	-3.1	-13.1	-26	0	-16	0.051	4,840
	L	Medial globus pallidus					-14	-8	2	0.019	
	L	Amygdala					-22	-14	-10	0.019	
	L	Lentiform nucleus					-8	0	0	0.017	
	L	Cingulate gyrus	24	0	11.9	32.7	-4	12	32	0.033	4,768
	R	Medial frontal gyrus	6				4	14	44	0.026	
	L	Cingulate gyrus	32				-2	22	26	0.024	
	L	Cingulate gyrus	24				-2	-2	30	0.023	
	L	Insula	13	-40.7	10.2	1.3	-40	10	2	0.031	3,312
	L	Clastrum					-34	10	6	0.030	
	L	Inferior frontal gyrus	47				-46	14	0	0.030	
HC > SP											
	R	Anterior cingulate	24	8.2	38.7	-1.2	8	38	-2	0.017	768
	L	Amygdala		-25	-2.7	-15.5	-24	0	-14	0.014	664
	L	Amygdala					-28	-8	-18	0.010	
	L	Anterior cingulate	32	-3.4	34.2	-9.6	-4	32	-6	0.011	456
	L	Anterior cingulate	32				-4	34	-10	0.011	
HC≠SP											
	L	Amygdala		-17.2	-5.5	-8.3	-24	0	-16	0.063	8,000
	L	Thalamus					-12	-8	2	0.030	
	L	Medial dorsal nucleus of thalamus					-2	-18	6	0.029	
	R	Ventral anterior nucleus of thalamus					14	-4	8	0.023	
	L	Amygdala					-22	-14	-10	0.019	
	R	Medial dorsal nucleus of thalamus					4	-14	10	0.018	
	R	Amygdala		28.2	2.1	-7.6	20	-4	-12	0.045	7,336
	R	Amygdala					26	0	-12	0.038	
	R	Insula	13				38	6	0	0.034	
	R	Inferior frontal gyrus	47				42	20	2	0.026	
	L	Cingulate gyrus	24	.2	11.1	33.5	-4	12	32	0.033	4,976

(Continues)

TABLE 2 (Continued)

Whole brain	Hemisphere	Region	BA	Center of mass			Peak			ALE value at peak	Volume (mm ³)
				X	Y	Z	X	Y	Z		
	R	Medial frontal Gyrus	6				4	14	44	0.026	
	R	Cingulate gyrus	24				4	2	42	0.025	
	L	Cingulate gyrus	32				-2	22	26	0.024	
	L	Cingulate gyrus	24				-2	-2	30	0.023	
	L	Insula	13	-40.4	10	1.2	-36	8	6	0.032	3,312
	L	Insula	13				-40	10	2	0.031	
	L	Inferior frontal gyrus	47				-46	14	0	0.030	
ROI	Hemisphere	Region	BA	Center of mass			Peak			ALE value at peak	Volume (mm ³)
SP > HC											
	R	Amygdala		29.2	3.6	-5.9	20	-4	-10	0.041	7,664
	R	Clastrum					38	6	2	0.030	
	R	Inferior frontal gyrus	45				44	22	4	0.024	
	L	Amygdala		-23.1	-2.6	-13.6	-26	-2	-16	0.050	5,344
	L	Parahippocampal gyrus	35				-22	-16	-10	0.018	
	L	Inferior frontal gyrus	47				-26	14	-12	0.016	
	L	Cingulate gyrus	24	-3	9.2	33.1	-2	8	32	0.027	3,376
	L	Cingulate gyrus	24				-2	0	30	0.022	
	R	Medial frontal gyrus	6				4	14	44	0.019	
	R	Cingulate gyrus	24				2	18	28	0.015	
	L	Insula	13	-40.1	10.4	2.8	-40	10	4	0.035	3,256
	L	Inferior frontal gyrus	47				-46	14	0	0.033	
HC > SP											
	R	Anterior cingulate	24	8.2	38.6	-1.3	8	38	-2	0.017	792
	L	Amygdala		-25.1	-2.5	-15.3	-24	0	-14	0.014	776
	L	Amygdala					-28	-8	-18	0.010	
HC ≠ SP											
	R	Amygdala		28.8	3.2	-6.6	20	-4	-12	0.043	7,704
	R	Clastrum					38	6	2	0.030	
	R	Inferior frontal gyrus	45				44	22	4	0.024	
	L	Amygdala		-22.9	-3.8	-12.6	-24	0	-16	0.061	6,992
	L	Medial globus pallidus					-14	-8	0	0.022	
	L	Parahippocampal gyrus	35				-22	-16	-10	0.018	
	L	Medial dorsal nucleus					-6	-12	8	0.016	
	L	Inferior frontal gyrus	47				-26	14	-12	0.016	
	L	Insula	13	-40	10.3	2.7	-40	10	4	0.035	3,248
	L	Inferior frontal gyrus	47				-46	14	0	0.033	
	L	Cingulate gyrus	24	-2	9.2	33	-2	8	32	0.027	3,096
	L	Cingulate gyrus	24				-2	0	30	0.022	
	R	Medial frontal gyrus	6				4	14	44	0.019	
	R	Cingulate gyrus	24				2	18	28	0.015	6

The table shows results for the whole brain analysis alone, combined whole brain analysis and ROI, and ROI alone.

first subsample of subjects took part in: Straube, Glauer, et al. (2006); Straube et al. (2004, 2007); Straube, Mentzel, et al. (2006). The second one was included in: Lipka et al. (2011, 2014); Straube et al. (2011), and the third subsample participated both in Hilbert et al. (2014) and Lueken et al. (2014).

Of note, Caseras and Lueken ensured that there was no overlap in participants who have been involved in: Caseras, Giampietro, et al. (2010); Caseras, Mataix-Cols, et al. (2010); Lueken et al. (2014); Lueken et al. (2011). All the studies in which an overlap of participants was present reported coordinates for the SP > HC contrast, whereas

TABLE 3 Summary of the ROI defined and analyzed in the original articles and selected for the present meta-analysis

		Considered as ROI (no. of articles)	Significantly recruited (no. of articles)	Significantly recruited in a single side (no. of articles)
ACC	L	20	11	4
	R	20	11	4
Amygdala	L	26	16	7
	R	26	12	3
Basal ganglia	L	2	1	0
	R	2	2	1
BNST	L	2	1	0
	R	2	2	1
Cuneus	L	1	1	0
	R	1	1	0
DLPFC	L	10	5	1
	R	10	4	2
DMPFC	L	9	5	1
	R	9	5	1
Fusiform	L	7	5	2
	R	7	4	1
Hippocampus	L	7	4	0
	R	7	5	1
IFG	L	2	2	0
	R	2	2	0
Insula	L	24	14	4
	R	24	15	5
IPFC	L	3	0	0
	R	3	0	0
IPL	L	1	0	0
	R	1	0	0
Midbrain	L	1	0	0
	R	1	1	1
MPFC	L	8	4	0
	R	8	5	1
Paracentral lobule	L	1	0	0
	R	1	0	0
Parahippocampus	L	2	1	0
	R	2	1	0
OFC	L	12	5	1
	R	12	5	1
PCC	L	1	1	1
	R	1	0	0
SMA	L	2	2	1
	R	2	1	0
SPFC	L	5	1	0
	R	5	1	0
SPL	L	1	0	0
	R	1	0	0
Thalamus	L	8	4	2
	R	8	3	1

ACC = anterior cingulate cortex; BNST = bed nucleus of the stria terminalis; DLPFC = dorsolateral prefrontal cortex; DMPFC = dorsomedial prefrontal cortex; IFG = inferior frontal gyrus; IPFC = inferior prefrontal cortex; IPL = inferior parietal lobule; MPFC = middle prefrontal cortex; OFC = orbitofrontal cortex; PCC = posterior cingulate cortex; SMA = supplementary motor area; SPFC = superior prefrontal cortex; SPL = superior parietal lobule.

the HC > SP was not affected by this issue. Therefore, we limited the two study overlap analyses to the SP > HC and to the omnibus contrast only (see Supporting Information Material for further details). The pooled-studies analysis yielded very small differences as compared with the original one, with an almost complete overlap among the significant clusters for all the different conditions (i.e., combined, ROI only and whole-brain only) and both contrasts (i.e., for SP > HC see Supporting Information Figure S5 and for SP ≠ HC see Supporting Information Figure S6). As expected, the leave-studies out analysis was more conservative and produced smaller clusters of differences between groups, even though their location and number was comparable with those obtained from the original analysis (i.e., for SP > HC see Supporting Information Figure S7 and for SP ≠ HC see Supporting Information Figure S8). Particularly, the voxelwise whole-brain results were less affected by this procedure, confirming the robustness of these findings (Figures 2c and 3c; Supporting Information Figures S5C, S6C, S7C, and S8C).

4 | DISCUSSION

4.1 | ALE meta-analysis results

The results of a voxelwise whole-brain meta-analysis, a ROI-based meta-analysis and their combination confirmed the consistent involvement of the *limbic circuit* in phobias. The right amygdala, the anterior portion of the midcingulate cortex, and the insula showed higher convergence of activation in SP patients as compared with controls during visual processing of phobic stimuli. All these regions are involved in fear and anxiety responses typically recruited while processing threatening and aversive stimuli (Chen, Wang, Wang, & Li, 2017; Tovote, Fadok, & Luthi, 2015). Of note, while our meta-analysis revealed a bilateral recruitment of the amygdala and insula, several studies reported asymmetric responses (e.g., Brinkmann et al., 2017; Britton et al., 2009; Lueken et al., 2014; Munsterkotter et al., 2015; Schienle et al., 2005; Schweckendiek et al., 2011; Straube, Mentzel, et al., 2006; Straube et al., 2007; Wiemer et al., 2015).

The inferior frontal gyrus (IFG) was found more active in individuals with phobias, and several studies contribute to the ALE cluster encompassing this region (Supporting Information Table S4). The pattern of activity of this brain area has been linked to different presentation modalities of phobic stimuli (e.g., words; Straube et al., 2004) and to different subtypes of phobias (e.g., spider phobia, blood, injury and injection phobia; Caseras, Giampietro, et al., 2010; Caseras, Mataix-Cols, et al., 2010; Michalowski et al., 2017). It has been suggested that an increment in hemodynamic activity in the IFG and, more generally, in the prefrontal cortex, may be related to the reduction of emotional appraisal abilities in patients with specific phobia (Caseras, Giampietro, et al., 2010; Michalowski et al., 2017; Straube et al., 2004). In addition, the ventral part of the cingulate cortex and the orbitofrontal cortex was modulated by emotional valence of the presented stimuli in several studies, both in normal and pathological conditions (Courtin, Bienvenu, Einarsson, & Herry, 2013; Del Casale et al., 2012; Drevets, Savitz, & Trimble, 2008; Potegal, 2012). Reduced activity in the medial

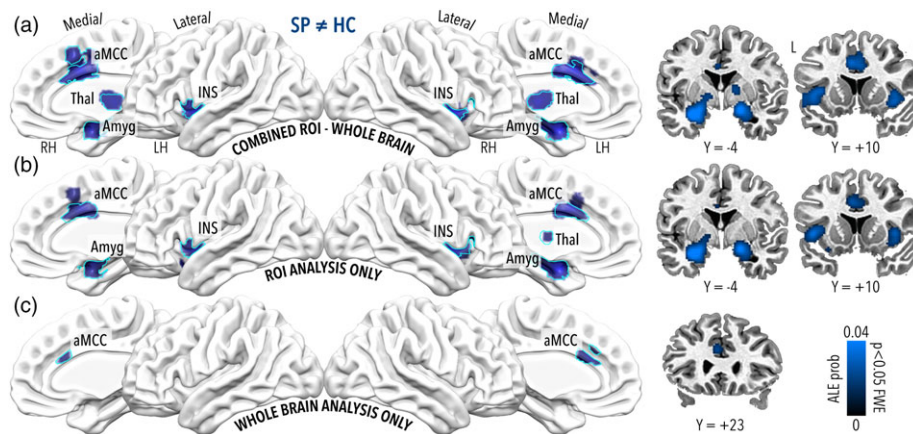


FIGURE 3 Significant results for the omnibus ALE meta-analysis contrast $SP \neq HC$ (ALE $p < 0.05$ corrected). Panel a describes results including both whole brain and ROI analysis, panel b clusters obtained from the ROI analysis only and panel c from the voxelwise whole brain analysis only. For each of the three conditions, results of the omnibus meta-analysis test revealed an almost perfect overlap with the pattern of significant clusters obtained from the $SP > HC$ contrast (Figure 2). The only relevant difference was found for the ROI only condition, where a cluster extending to the left thalamus was present in the omnibus test (panel b) but not in the $SP > HC$ contrast (Figure 2). Cyan outline represents the results for the leave-studies out analysis of the omnibus $SP \neq HC$ contrast of interest, as described in the *Study overlap analysis* paragraph of the *Methods* section [Color figure can be viewed at wileyonlinelibrary.com]

prefrontal cortex may probably reflects a phobia-specific down-regulation deficit in patients (Del Casale et al., 2012).

Our results showed altered activity in the left amygdala for both the $HC > SP$ and the $SP > HC$ contrasts. This evidence indicates that when this region is considered and a high sensitivity method (i.e., ROI approach) is adopted, differences between the two groups may arise. However, whether the left amygdala is more (or less) recruited in HC as compared with SP may depend on intervening variables (e.g., experimental paradigm, stimuli, and study design) and/or to inconsistencies among different studies/experiments captured by meta-analyses.

In addition, of the three studies reporting greater activations of the left amygdala for controls as compared with SP patients (one in the supplementary materials and the other two in the main manuscript; Killgore et al., 2014; Schweckendiek et al., 2011; Straube, Glauer, et al., 2006; Supporting Information Table S7), none of them discussed this finding. This issue may arise when researchers focus on confirming the primary hypothesis of the study, failing to describe counterintuitive results that may, even partially, affect the original hypothesis. Nonetheless, further studies are needed to evaluate the consistency of this finding.

Importantly, considering studies based on voxelwise whole-brain approaches exclusively, the only significant finding was an increased activation in the anterior portion of the midcingulate cortex in individuals with phobias as compared with controls.

Regarding the possible biases that may influence meta-analysis results, it is interesting to note that our RoB evaluation, based on a modified version of the NOS, highlighted that less than half of the considered studies can be classified as low risk. We believe that the assessment of study quality based on the neuroimaging-adapted NOS scale represents a crucial point to ensure the robustness of meta-analysis results. Therefore, we hope that this, or other similar procedures, would become a standard in future neuroimaging meta-analyses and systematic reviews.

4.2 | Analysis of ROIs occurrence

We also performed an estimate of the occurrence of ROIs in all the selected studies. As expected, regions pertaining to the *limbic circuit* not only occurred more frequently, but in most of these cases activation was significantly different between the two groups. However, when negative results were found, they were scarcely discussed. Specifically, 10 studies focusing on the amygdala reported unilateral activations without discussing this unexpected result. In addition, seven studies investigating the same region found no significant differences between individuals with phobias and controls. Conversely, other brain areas, including prefrontal, parietal and occipital visual cortices led to inconsistent results, although they were frequently studied. For instance, the orbito-frontal cortex was included as ROI in 12 studies, but it was significantly more active in individuals with phobias only in 6 of them, and in only 4 of these with significant bilateral activation. These results suggest that the definition of some ROIs did not rely on strong or consistent a priori hypothesis. We would argue that the systematic presence of negative results would foster the revision of theories used to derive a priori ROI hypotheses in the first place.

4.3 | On the use of ROIs in specific phobias fMRI studies

Neuroimaging studies on phobias make large use of the ROI approach. In our meta-analysis, less than 10% (3 out of 31) of all the included studies used a whole-brain analysis approach exclusively, whereas about 22% of the studies (7 out of 31) employed the ROI approach only. With respect to the rest of the studies in which both the approaches were used, the whole-brain analysis has been considered just as an exploratory analysis. A closer look at studies based on ROI analysis, demonstrated that a variety of methods is employed to define regions boundaries. Also, in the rationale for choosing a particular region some heterogeneity is present: in our search 8 out of

27 studies did not report accurate literature references and 5 did not even report a rationale for the selection (Supporting Information Table S9). The lack of consensus in ROI selection and definition likely affects reproducibility and generalizability of the results.

The comparison of results obtained from nine meta-analyses based on different methodological pipelines shows that important differences can arise depending on the inclusion of ROI-based studies. ROI analysis is a widely used neuroimaging method that aims at testing specific a priori hypotheses and, compared with voxelwise whole-brain analysis, has the advantage of reducing the number of tests performed and mitigating the multiple comparisons problem. Issues surrounding multiple comparisons represent a serious problem when performing mass-univariate statistics, limiting the chance of getting true significant results, especially when small samples are considered (Poldrack et al., 2017; Yarkoni, 2009). Although the ROI approach has been considered as a valid alternative to voxelwise statistics, additional caveats need to be considered. In functional neuroimaging studies, the ROI selection process is often based on anatomical landmarks; however, this may not be the optimal choice because there is no evidence that the overlap between brain function and structure is consistent, or even constant throughout the brain (Huettel et al., 2004). In addition, even when a functional localizer experiment defines ROI boundaries, there is no guarantee that the characterization of functional anatomy is not biased (Friston, Rotshtein, Geng, Sterzer, & Henson, 2006), and a small proportion of mental functions can be precisely and unequivocally localized (Poldrack, 2011). Another concern is related to which measure best represents the ROI signal: typically, the average signal across voxels is used. However, when only a small proportion of voxels are active within a large region, the average activity would reduce one's ability to detect significant effects (Poldrack, 2007). Furthermore, a judicious theoretical rationale should support the a priori identification of ROIs. In this regard, the lack of significant results in ROI analysis should force researchers to reconsider their own theories and update the estimates of whether and to what extent a brain region is involved in a specific mental process. On the contrary, relegating such results to the supplementary information or, even worse, neglecting them completely, increases the risk of confirmation bias and of distorting the neuroscientific literature.

A second order issue is related to the inclusion of studies based on the ROI method in neuroimaging meta-analysis. Recent guidelines report that ROI studies should not be included in ALE meta-analysis, since the "*inclusion of heterogeneous region-of-interest (ROI) or small volume corrected (SVC) analyses would violate this assumption and lead to inflated significance for those regions that come from overrepresented ROI/SVC analyses*" (Müller et al., 2018; p. 154). Our findings, perhaps not surprisingly, offer empirical support to this recommendation, by showing that the inclusion of ROI-based studies led to substantially different results, as compared with those obtained from studies relying on the whole-brain analysis.

This discrepancy calls for a discussion on how to summarize different methodological approaches used in the fMRI literature. Indeed, researchers typically take advantage of the ROI-based approach in many fields of social, cognitive and affective neuroscience; therefore, ignoring these types of studies may also lead to biased meta-analytic results as well. For instance, the ROI approach is sometimes employed

for questions related to smaller and anatomically well-defined regions (e.g., amygdala), so as to circumvent the intrinsic disadvantage introduced by the minimum cluster size when cluster-based correction methods are used. This is particularly true since in a vast amount of mass-univariate whole-brain fMRI studies, researchers opt for cluster-based correction methods to control for false positive results. For instance, in our meta-analysis, 21 studies over 31 applied some form of cluster-based correction. It should be noted that canonical cluster-based correction methods could theoretically capture activity of smaller brain regions when the cluster-forming threshold is sufficiently stringent, since the higher this value is, the smaller is the minimum cluster size threshold for significance (Cox et al., 2016; Eklund et al., 2016). However, in our sample, only 5 out of 31 studies employed such a high threshold. Thus, for whole-brain ALE meta-analysis maps, the lack of significant results for smaller brain regions, as the amygdala in our case, may equally reflect a true negative or a false negative result, the latter being a mere consequence of thresholding. One option to overcome this problem would be the use of correction methods that weigh both the extent and intensity of the activation and do not require specifying a single value for the cluster-forming threshold, as the threshold-free cluster enhancement algorithm (Smith & Nichols, 2009). These methods may increase the sensitivity of ALE maps in capturing true effects even in smaller brain regions. In addition, considering meta-studies some authors recently provided interesting analysis tools aimed at solving the issue of different thresholding methods (Costafreda, 2009, 2012) although there is still no extensive testing of such techniques in literature. Moreover, the sharing of unthresholded maps of brain activity (e.g., <https://neurovault.org/>; Gorgolewski et al., 2015) will foster the growth of meta-analysis studies that would not rely on coordinates of statistical significance and would not be affected by the aforementioned problems. In particular, the availability of unthresholded maps would, at least, prevent problems related to differences in thresholding and in multiple comparisons correction methods across studies (Carp, 2012; Eklund et al., 2016). Image-based meta-analysis including large samples of unthresholded maps would represent for neuroimaging what individual patient meta-analysis represents for traditional meta-analysis (Riley, Lambert, & Abo-Zaid, 2010). In addition, the use of unthresholded maps would mitigate the effects of different statistical procedures adopted across studies to overcome the multiple comparisons issue. Indeed, this heterogeneity of methods likely affects the results obtained from meta-analyses, as this information cannot be included in the standard ALE pipeline. Further, a very large number of studies would be needed to estimate the impact of different statistical procedures and this may be hardly achieved in the domain of specific phobias, but also considering other more extensively studied topics.

The overcoming of ROI-related issues may be also achieved by radically changing the way in which researchers acquire and analyze neuroimaging data, as for instance in the Human Connectome Project (Coalson, Van Essen, & Glasser, 2018; Glasser, Glasser, & Smith, 2016). As a matter of fact, high-quality multimodal datasets can help in the precise definition of anatomo-functional ROI at single-subject level, as compared with the definition of parcels based on standard-space normalization. This would ultimately lead to the definition of ROIs in which the anatomical and functional specificity is preserved,

so that brain parcellation is based on characteristics that can be generalized (e.g., myelination, thickness, resting state activity) rather than determined ad-hoc in every single study.

In the present study, we assessed how frequently certain brain regions were considered of interest across the selected studies, and by simply counting the reports where each region was found to be significantly modulated by SP, we clearly highlighted the discrepancy between what researchers hypothesized and the actual results. Although this is a simple, descriptive and semi-quantitative method to summarize findings coming from different ROI studies, it could provide some relevant insights, as also suggested by others (Müller et al., 2018).

5 | LIMITATIONS

In the present study, we did not contrast different subtypes of phobias, especially since there are very few studies that performed this direct comparison (Caseras, Giampietro, et al., 2010; Caseras, Mataix-Cols, et al., 2010; Lueken et al., 2011, 2014). This question is particularly relevant for some specific phobias like blood, injury and injection, characterized by specific behavioral and neurovegetative responses as compared with other subtypes (e.g., spider phobias or small animal phobias; Sarlo, Buodo, Munafò, Stegagno, & Palomba, 2008; Sarlo, Palomba, Angrilli, & Stegagno, 2002). A possible strategy to overcome this limitation is to first compute meta-analytic maps for studies including each single group of patients (i.e., SP alone, control alone etc.) and then to perform a second level meta-analytic contrast. Nonetheless, given the reduced amount of studies reporting these activations, we did not perform this analysis, also because this aspect falls beyond the scopes of the present work.

A second limitation is related to the results for the HC > SP comparison, which have been reported in few studies. When we personally contacted corresponding authors of included researches, they confirmed that lack of significant results was not related to omissions in the reports, but reflected the fact that the HC > SP contrast did not reach the statistical significance level. For this reason, as pointed out in Müller et al. (2018), the number of studies reporting results for this comparison may not be sufficient to ensure the validity of current findings. Importantly, our further investigation regarding the HC > SP contrast also showed that negative results, especially when related to contrasts that are not the main focus of interest of the study, are seldom or not reported at all.

Lastly, considering RoB, it was not possible to perform sensitivity analysis (excluding low and medium risk studies) because of the small number of included investigations (Eickhoff, 2016; Müller et al., 2018).

6 | CONCLUSIONS

With over 90% of the studies considered in our meta-analysis including the ROI approach (alone or in combination with whole-brain analysis), it is evident how results obtained with this method deeply influence our current understanding of neurobiology of phobias. This may raise possible concerns: with the inclusion of ROI-based studies, results showed increased convergence of activations in the right

amygdala, the insula, and the cingulate cortex of phobic patients as compared with controls. Importantly, anterior portion of the midcingulate cortex is the only region that survived when we subsequently considered studies based on voxelwise whole-brain analysis only. This result highlights the impact of ROI-based findings in neuroimaging meta-analysis. In this regard, a focus on replication studies with larger number of participants is critical and this will also inherently provide researchers the statistical power required to test their hypotheses using a more data-driven approach, such as the voxelwise whole-brain one.

Furthermore, particular care should be used when considering ROI studies, since many of them lack a strong a priori hypothesis in the definition and selection of brain areas. Moreover, several selected ROIs did not reach the statistical significance for between-groups comparisons and negative results were typically relegated to supplementary sections or completely neglected. Interestingly, significant results in contrast with the original hypothesis were also not taken into account in the manuscript (e.g., the higher activation of a portion of the amygdala in controls as compared with individuals with phobias). Overall, these findings suggest that ROIs should be used judiciously and parsimoniously and restricted to cases where a strong a priori hypothesis is formulated. All negative and contradictory results should be discussed, so that researchers could revise their initial hypotheses based on actual results.

ACKNOWLEDGMENTS

The authors wish to thank the following researchers for their help during the reviewing phases of the present manuscript, providing relevant information about possible overlaps of participants among different published studies and about possible missing information regarding significant clusters of activations: Dr. Thomas Straube and Dr. Udo Dannlowski (Muenster University); Dr. Kevin Hilbert and Dr. Ulrike Leuken (Humboldt-Universität zu Berlin); Dr. Paul Pauli and Dr. Julian Wiemer (Würzburg University); Dr. Zilverstand (Icahn School of Medicine at Mount Sinai), Dr. Ulrike Halsband (Freiburg University), Dr. Xavier Caseras (University of Cardiff). The authors would also especially thank Dr. Hilbert and Lueken for providing data from an errata they are writing regarding the article *Fear Processing in Dental Phobia during Crossmodal Symptom Provocation: An fMRI Study - BioMed Research International 2014*.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

ORCID

Claudio Gentili  <https://orcid.org/0000-0002-2579-8755>

Simone Messerotti Benvenuti  <https://orcid.org/0000-0002-1430-6807>

Giada Lettieri  <https://orcid.org/0000-0001-8027-9053>

Cristiano Costa  <https://orcid.org/0000-0001-5472-546X>

Luca Cecchetti  <https://orcid.org/0000-0001-5184-6477>

REFERENCES

- APA. (2013). *Diagnostic and statistical manual of mental disorders : DSM-5*. Washington, DC: American Psychiatric Association.
- Brinkmann, L., Poller, H., Herrmann, M. J., Miltner, W., & Straube, T. (2017). Initial and sustained brain responses to threat anticipation in blood-injection-injury phobia. *NeuroImage: Clinical*, *13*, 320–329.
- Britton, J. C., Gold, A. L., Deckersbach, T., & Rauch, S. L. (2009). Functional MRI study of specific animal phobia using an event-related emotional counting stroop paradigm. *Depression and Anxiety*, *26*, 796–805.
- Carp, J. (2012). On the plurality of (methodological) worlds: Estimating the analytic flexibility of fMRI experiments. *Frontiers in Neuroscience*, *6*, 149.
- Caseras, X., Giampietro, V., Lamas, A., Brammer, M., Vilarroya, O., Carmona, S., ... Mataix-Cols, D. (2010). The functional neuroanatomy of blood-injection-injury phobia: A comparison with spider phobics and healthy controls. *Psychological Medicine*, *40*, 125–134.
- Caseras, X., Mataix-Cols, D., Trasovares, M. V., Lopez-Sola, M., Ortriz, H., Pujol, J., ... Torrubia, R. (2010). Dynamics of brain responses to phobic-related stimulation in specific phobia subtypes. *The European Journal of Neuroscience*, *32*, 1414–1422.
- Chen, W., Wang, Y., Wang, X., & Li, H. (2017). Neural circuits involved in the renewal of extinguished fear. *IUBMB Life*, *69*, 470–478.
- Coalson, T. S., Van Essen, D. C., & Glasser, M. F. (2018). The impact of traditional neuroimaging methods on the spatial localization of cortical areas. *Proceedings of the National Academy of Sciences*, *115*, E6356–E6365.
- Costafreda, S. G. (2009). Pooling fMRI data: Meta-analysis, mega-analysis and multi-center studies. *Frontiers in Neuroinformatics*, *3*, 33.
- Costafreda, S. G. (2012). Parametric coordinate-based meta-analysis: Valid effect size meta-analysis of studies with differing statistical thresholds. *Journal of Neuroscience Methods*, *210*, 291–300.
- Courtin, J., Bienvenu, T. C., Einarsson, E. O., & Herry, C. (2013). Medial prefrontal cortex neuronal circuits in fear behavior. *Neuroscience*, *240*, 219–242.
- Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, *29*, 162–173.
- Cox, R. W., Reynolds, R. C., & Taylor, P. A. (2016). AFNI and clustering: False positive rates redux. *bioRxiv*, 2016, 065862.
- Cremers, H. R., Wager, T. D., & Yarkoni, T. (2017). The relation between statistical power and inference in fMRI. *PLoS One*, *12*, e0184923.
- Del Casale, A., Ferracuti, S., Rapinesi, C., Serata, D., Piccirilli, M., Savoia, V., ... Girardi, P. (2012). Functional neuroimaging in specific phobia. *Psychiatry Research*, *202*, 181–197.
- Dilger, S., Straube, T., Mentzel, H.-J., Fitzek, C., Reichenbach, J. R., Hecht, H., ... Miltner, W. H. R. (2003). Brain activation to phobia-related pictures in spider phobic humans: An event-related functional magnetic resonance imaging study. *Neuroscience Letters*, *348*, 29–32.
- Drevets, W. C., Savitz, J., & Trimble, M. (2008). The subgenual anterior cingulate cortex in mood disorders. *CNS Spectrums*, *13*, 663–681.
- Eickhoff, S. B. (2016). Behavior, sensitivity, and power of activation likelihood estimation characterized by massive empirical simulation. *NeuroImage*, *137*, 70–85.
- Eickhoff, S. B., Laird, A. R., Grefkes, C., Wang, L. E., Zilles, K., & Fox, P. T. (2009). Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: A random-effects approach based on empirical estimates of spatial uncertainty. *Human Brain Mapping*, *30*, 2907–2926.
- Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 7900–7905.
- Friston, K. J., Rotshtein, P., Geng, J. J., Sterzer, P., & Henson, R. N. (2006). A critique of functional localisers. *NeuroImage*, *30*, 1077–1087.
- Glasser, M. F., Glasser, M. F., & Smith, S. M. (2016). The human connectome project's neuroimaging approach. *Proceedings of the National Academy of Sciences of the United States of America*, *119*, 1175–1187.
- Goossens, L., Schruers, K., Peeters, R., Griez, E., & Snaert, S. (2007a). Visual presentation of phobic stimuli: Amygdala activation via an extrageniculostriate pathway? *Psychiatry Research*, *155*, 113–120.
- Goossens, L., Snaert, S., Peeters, R., Griez, E. J., & Schruers, K. R. (2007b). Amygdala hyperfunction in phobic fear normalizes after exposure. *Biological Psychiatry*, *62*, 1119–1125.
- Gorgolewski, K. J., Varoquaux, G., Rivera, G., Schwarz, Y., Ghosh, S. S., Maumet, C., ... 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*, 8.
- Halsband, U., & Wolf, T. G. (2015). Functional changes in brain activity after hypnosis in patients with dental phobia. *Journal of Physiology, Paris*, *109*, 131–142.
- Hermann, A., Schafer, A., Walter, B., Stark, R., Vaitl, D., & Schienle, A. (2007). Diminished medial prefrontal cortex activity in blood-injection-injury phobia. *Biological Psychology*, *75*, 124–130.
- Hilbert, K., Evens, R., Maslowski, N. I., Wittchen, H. U., & Lueken, U. (2014). Fear processing in dental phobia during crossmodal symptom provocation: An fMRI study. *BioMed Research International*, 2014, 196353.
- Huettel, S. A., Song, A. W., & McCarthy, G. (2004). *Functional magnetic resonance imaging*. Sunderland: Sinauer Associates.
- Killgore, W. D., Britton, J. C., Schwab, Z. J., Price, L. M., Weiner, M. R., Gold, A. L., ... Rauch, S. L. (2014). Cortico-limbic responses to masked affective faces across PTSD, panic disorder, and specific phobia. *Depression and Anxiety*, *31*, 150–159.
- Lieberman, M. D., & Cunningham, W. A. (2009). Type I and type II error concerns in fMRI research: Re-balancing the scale. *Social Cognitive and Affective Neuroscience*, *4*, 423–428.
- Lipka, J., Hoffmann, M., Miltner, W. H., & Straube, T. (2014). Effects of cognitive-behavioral therapy on brain responses to subliminal and supraliminal threat and their functional significance in specific phobia. *Biological Psychiatry*, *76*, 869–877.
- Lipka, J., Miltner, W. H., & Straube, T. (2011). Vigilance for threat interacts with amygdala responses to subliminal threat cues in specific phobia. *Biological Psychiatry*, *70*, 472–478.
- Lueken, U., Hilbert, K., Stolyar, V., Maslowski, N. I., Beesdo-Baum, K., & Wittchen, H. U. (2014). Neural substrates of defensive reactivity in two subtypes of specific phobia. *Social Cognitive and Affective Neuroscience*, *9*, 1668–1675.
- Lueken, U., Kruschwitz, J. D., Muehlhan, M., Siebert, J., Hoyer, J., & Wittchen, H. U. (2011). How specific is specific phobia? Different neural response patterns in two subtypes of specific phobia. *NeuroImage*, *56*, 363–372.
- Michalowski, J. M., Matuszewski, J., Drozdziel, D., Koziejowski, W., Rynkiewicz, A., Jednorog, K., & Marchewka, A. (2017). Neural response patterns in spider, blood-injection-injury and social fearful individuals: New insights from a simultaneous EEG/ECG-fMRI study. *Brain Imaging and Behavior*, *11*, 829–845.
- Müller, V. I., Cieslik, E. C., Laird, A. R., Fox, P. T., Radua, J., Mataix-Cols, D., ... Eickhoff, S. B. (2018). Ten simple rules for neuroimaging meta-analysis. *Neuroscience & Biobehavioral Reviews*, *84*, 151–161.
- Munsterkotter, A. L., Notzon, S., Redlich, R., Grotegerd, D., Dohm, K., Arolt, V., ... Dannlowski, U. (2015). Spider or no spider? Neural correlates of sustained and phasic fear in spider phobia. *Depression and Anxiety*, *32*, 656–663.
- Poldrack, R. A. (2007). Region of interest analysis for fMRI. *Social Cognitive and Affective Neuroscience*, *2*, 67–70.
- Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: From reverse inference to large-scale decoding. *Neuron*, *72*, 692–697.
- Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò, M. R., ... Yarkoni, T. (2017). Scanning the horizon: Towards transparent and reproducible neuroimaging research. *Nature Reviews. Neuroscience*, *18*, 115–126.
- Potegal, M. (2012). Temporal and frontal lobe initiation and regulation of the top-down escalation of anger and aggression. *Behavioural Brain Research*, *231*, 386–395.
- Riley, R. D., Lambert, P. C., & Abo-Zaid, G. (2010). Meta-analysis of individual participant data: Rationale, conduct and reporting. *BMJ*, *340*, 521–525.
- Sarlo, M., Buodo, G., Munafò, M., Stegagno, L., & Palomba, D. (2008). Cardiovascular dynamics in blood phobia: Evidence for a key role of sympathetic activity in vulnerability to syncope. *Psychophysiology*, *45*, 1038–1045.

- Sarlo, M., Palomba, D., Angrilli, A., & Stegagno, L. (2002). Blood phobia and spider phobia: Two specific phobias with different autonomic cardiac modulations. *Biological Psychology*, *60*, 91–108.
- Scharmuller, W., Ubel, S., Leutgeb, V., Schoengassner, F., Wabnegger, A., & Schienle, A. (2014). Do not think about pain: Neural correlates of attention guiding during visual symptom provocation in dental phobia—an fMRI study. *Brain Research*, *1566*, 69–76.
- Schienle, A., Schafer, A., Hermann, A., Rohrmann, S., & Vaitl, D. (2007). Symptom provocation and reduction in patients suffering from spider phobia: An fMRI study on exposure therapy. *European Archives of Psychiatry and Clinical Neuroscience*, *257*, 486–493.
- Schienle, A., Schafer, A., Walter, B., Stark, R., & Vaitl, D. (2005). Brain activation of spider phobics towards disorder-relevant, generally disgust- and fear-inducing pictures. *Neuroscience Letters*, *388*, 1–6.
- Schienle, A., Scharmuller, W., Leutgeb, V., Schafer, A., & Stark, R. (2013). Sex differences in the functional and structural neuroanatomy of dental phobia. *Brain Structure & Function*, *218*, 779–787.
- Schweckendiek, J., Klucken, T., Merz, C. J., Tabbert, K., Walter, B., Ambach, W., ... Stark, R. (2011). Weaving the (neuronal) web: Fear learning in spider phobia. *NeuroImage*, *54*, 681–688.
- Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, *44*, 83–98.
- Stang, A. (2010). Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. *European Journal of Epidemiology*, *25*, 603–605.
- Straube, T., Glauer, M., Dilger, S., Mentzel, H. J., & Miltner, W. H. (2006). Effects of cognitive-behavioral therapy on brain activation in specific phobia. *NeuroImage*, *29*, 125–135.
- Straube, T., Lipka, J., Sauer, A., Mothes-Lasch, M., & Miltner, W. H. (2011). Amygdala activation to threat under attentional load in individuals with anxiety disorder. *Biology of Mood & Anxiety Disorders*, *1*, 12.
- Straube, T., Mentzel, H. J., Glauer, M., & Miltner, W. H. (2004). Brain activation to phobia-related words in phobic subjects. *Neuroscience Letters*, *372*, 204–208.
- Straube, T., Mentzel, H. J., & Miltner, W. H. (2006). Neural mechanisms of automatic and direct processing of phobogenic stimuli in specific phobia. *Biological Psychiatry*, *59*, 162–170.
- Straube, T., Mentzel, H. J., & Miltner, W. H. (2007). Waiting for spiders: Brain activation during anticipatory anxiety in spider phobics. *NeuroImage*, *37*, 1427–1436.
- Tovote, P., Fadok, J. P., & Luthi, A. (2015). Neuronal circuits for fear and anxiety. *Nature Reviews Neuroscience*, *16*, 317–331.
- Turkeltaub, P. E., Eickhoff, S. B., Laird, A. R., Fox, M., Wiener, M., & Fox, P. (2012). Minimizing within-experiment and within-group effects in activation likelihood estimation meta-analyses. *Human Brain Mapping*, *33*, 1–13.
- Wendt, J., Lotze, M., Weike, A. I., Hosten, N., & Hamm, A. O. (2008). Brain activation and defensive response mobilization during sustained exposure to phobia-related and other affective pictures in spider phobia. *Psychophysiology*, *45*, 205–215.
- Wiener, J., Schulz, S. M., Reicherts, P., Glotzbach-Schoon, E., Andreatta, M., & Pauli, P. (2015). Brain activity associated with illusory correlations in animal phobia. *Social Cognitive and Affective Neuroscience*, *10*, 969–977.
- Wright, C. I., Martis, B., McMullin, K., Shin, L. M., & Rauch, S. L. (2003). Amygdala and insular responses to emotionally valenced human faces in small animal specific phobia. *Biological Psychiatry*, *54*, 1067–1076.
- Yarkoni, T. (2009). Big correlations in little studies: Inflated fMRI correlations reflect low statistical power—commentary on Vul et al. (2009). *Perspectives on Psychological Science*, *4*, 294–298.
- Zilverstand, A., Sorger, B., Kaemingk, A., & Goebel, R. (2017). Quantitative representations of an exaggerated anxiety response in the brain of female spider phobics—a parametric fMRI study. *Human Brain Mapping*, *38*, 3025–3038.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Gentili C, Messerotti Benvenuti S, Lettieri G, Costa C, Cecchetti L. ROI and phobias: The effect of ROI approach on an ALE meta-analysis of specific phobias. *Hum Brain Mapp*. 2019;40:1814–1828. <https://doi.org/10.1002/hbm.24492>