Nature Methods

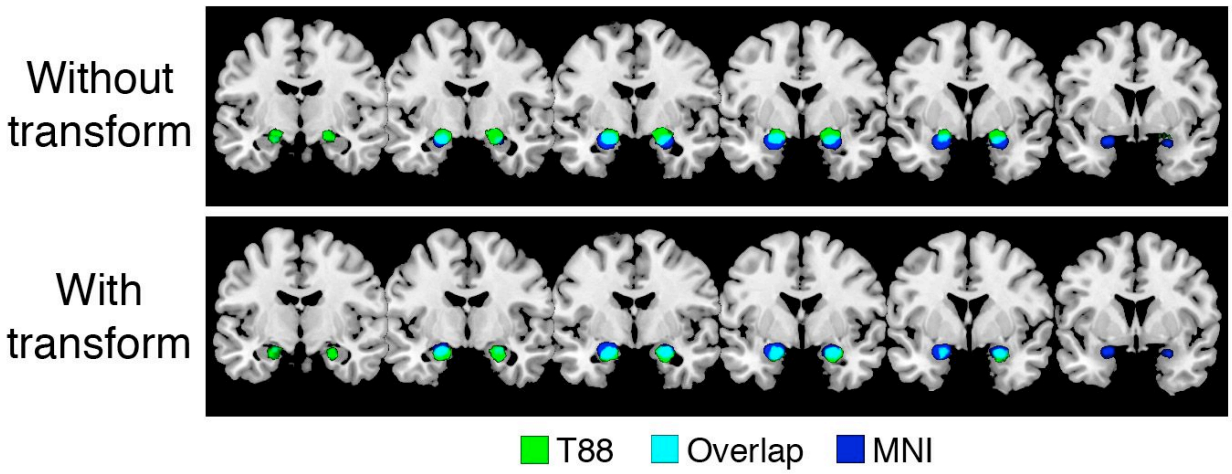
Large-scale automated synthesis of human functional neuroimaging data

Tal Yarkoni, Russell A Poldrack, Thomas E Nichols, David C Van Essen & Tor D Wager

Supplementary Figure 1	Effects of automated stereotactic space detection and correction on meta-analyses for the term 'amygdala'.
Supplementary Figure 2	Comparison of selected automatically-defined regions with anatomical regions delineated in the Harvard-Oxford Atlas.
Supplementary Figure 3	Category-specific activations in inferotemporal cortex identified with *NeuroSynth*.
Supplementary Figure 4	Whole-brain reverse inference maps for title-only analysis.
Supplementary Figure 5	Correlogram displaying pair-wise similarities between manually-generated meta-analysis and mega-analysis maps.
Supplementary Figure 6	Overlap between meta-analyses based on automatically-coded vs. manually-coded pain data.
Supplementary Figure 7	Whole-brain meta-analytic posterior probability maps for 25 key terms that occurred at high frequency (> 1 in 1,000 words) in at least 100 different studies in our database
Supplementary Figure 8	Mean probability of activation at each brain voxel across all 3,489 studies in the database.
Supplementary Figure 9	Accuracy of the naive Bayes classifier as a function of the number of terms being classified.
Supplementary Figure 10	Histograms showing frequency distributions of (A) proportion of voxels reported active in each study and (B) proportion of all studies reporting activity at each voxel
Supplementary Note	Additional validation analyses

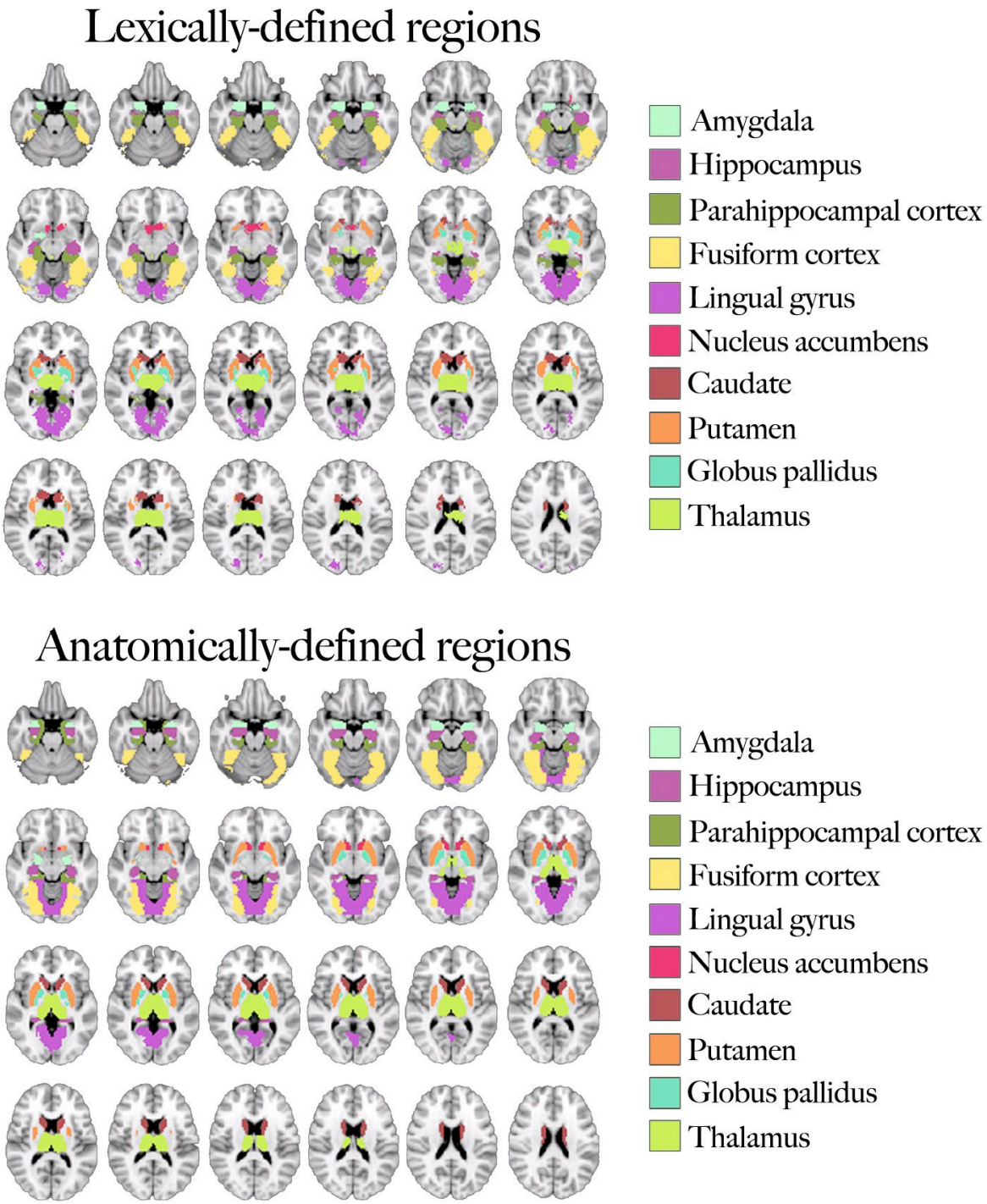
Supplementary Figures

Supplementary Figure 1



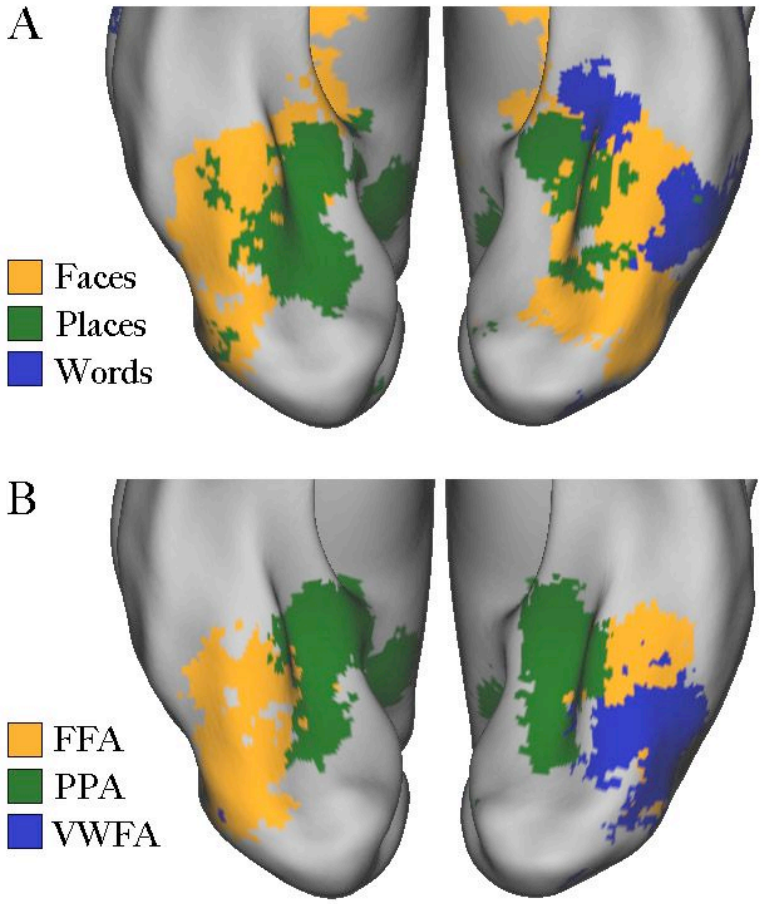
Effects of automated stereotactic space detection and correction on meta-analyses for the term 'amygdala'. Blue: regions maximally active in an automated meta-analysis of all studies that reported coordinates in MNI space. Green: regions maximally active when analyzing all studies reporting coordinates in T88 (i.e., Talairach-Tournoux, 1988[1]). Top row: overlap between MNI and T88 results space prior to application of any transformation. Notice the relatively poor alignment of the T88-based results with both the anatomical underlay and the MNI-based functional results (T88/MNI overlap colored cyan). Bottom row: following automated application of the Lancaster et al transform[2], the T88-based results are substantially better aligned along the dorsal/ventral axis, though differences remain along the rostral/caudal axis. The pearson correlation between MNI and T88 maps across all voxels improved from 0.73 pre-transform to 0.81 post-transform.

Supplementary Figure 2



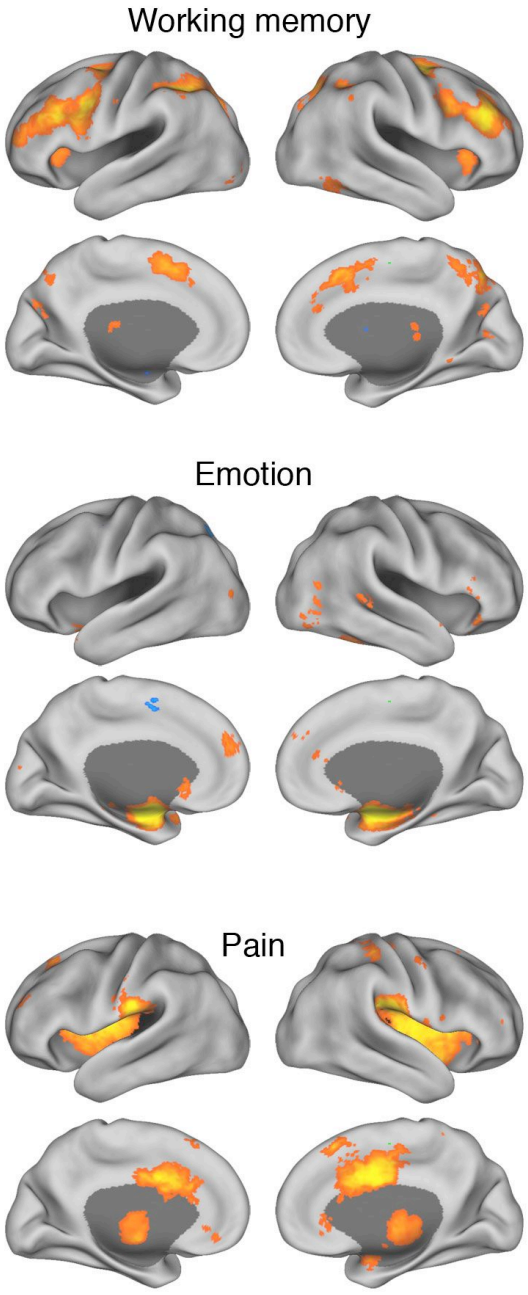
Comparison of selected automatically-defined regions with anatomical regions delineated in the Harvard-Oxford Atlas. Because the posterior probability maps could be thresholded arbitrarily, we restricted each automatically-generated ROI to the same number of voxels present in the ROI defined in the Harvard-Oxford atlas.

Supplementary Figure 3



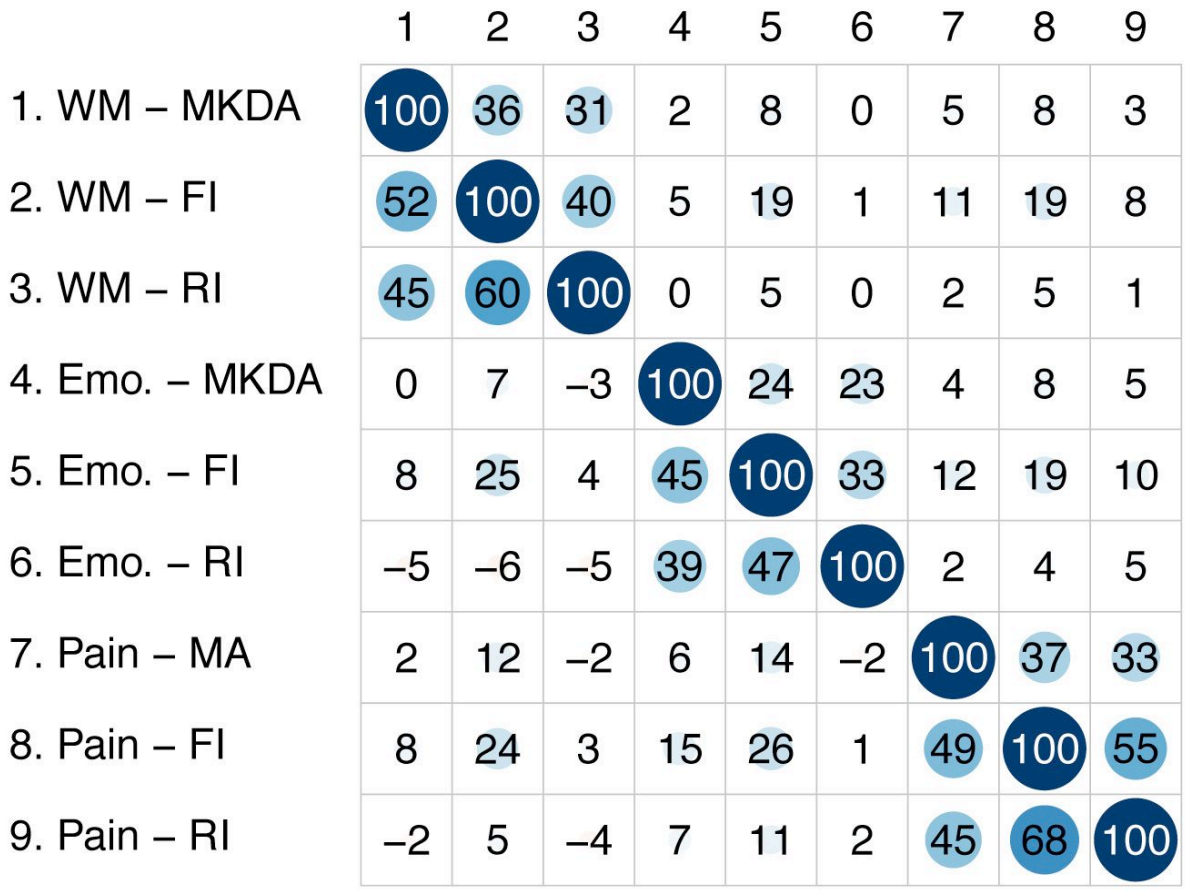
Category-specific activations in inferotemporal cortex identified with *NeuroSynth*. (A) Classification based on functional terms ('faces', 'places', and 'words'). (B) Classification based on putative functionally specialized cortical regions ('FFA', 'PPA', and 'VWFA'). Each voxel was assigned to the class with the highest posterior probability given observed activation at that voxel (e.g., in (A), observing activation in green voxels would imply a higher probability that the study was about places rather than faces or words).

Supplementary Figure 4



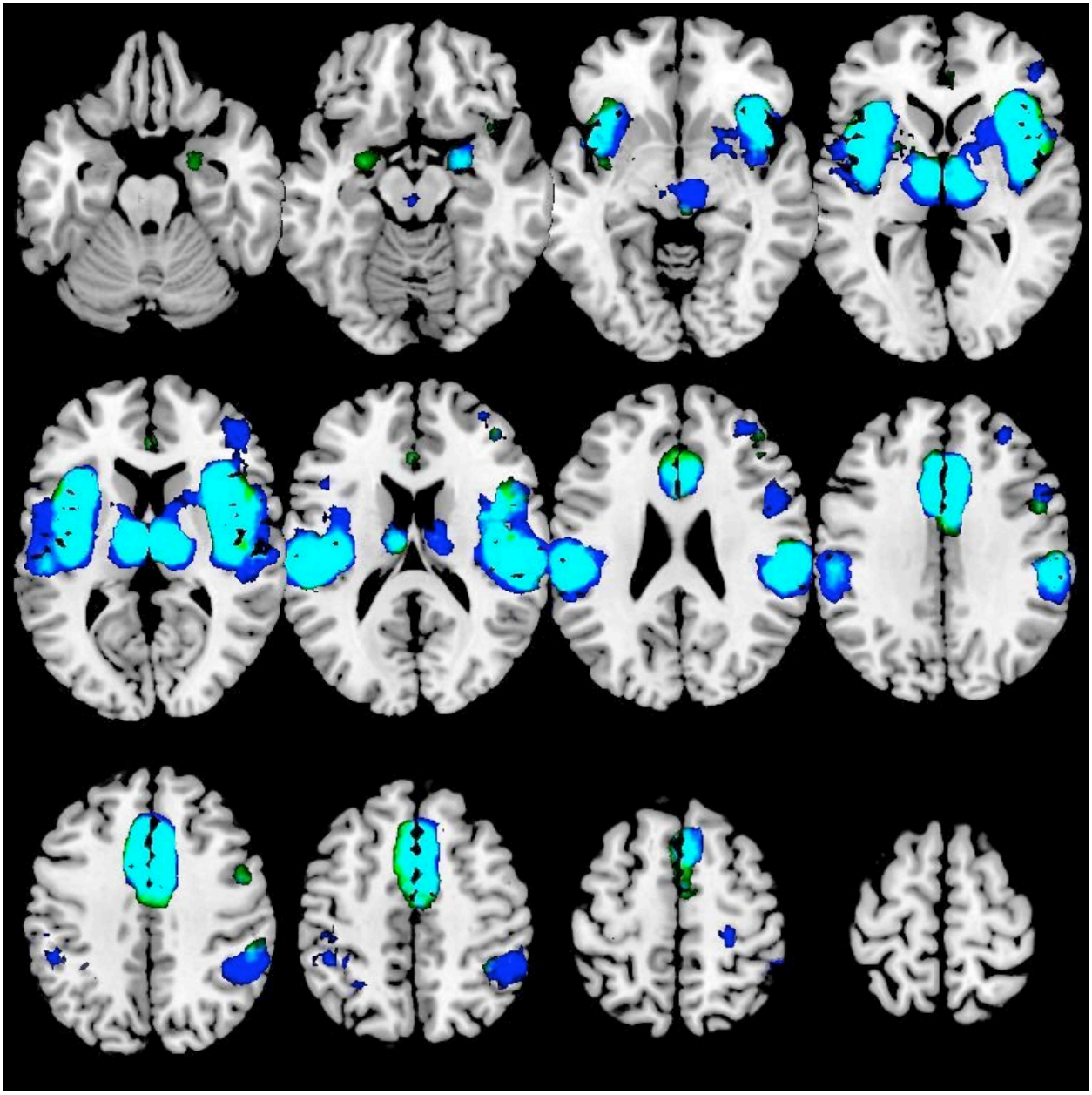
Whole-brain reverse inference meta-analysis maps for the terms 'working memory', 'emotion', and 'pain' when articles are coded strictly based on occurrence of terms in article title rather than in the full article text. Note the close similarity (along with apparent decreased sensitivity) relative to the full analyses reported in the text (Figure 2), suggesting that different sections of published articles carry broadly similar information.

Supplementary Figure 5



Correlogram displaying pair-wise similarities between manually-generated meta-analysis and mega-analysis maps (Fig. 2A) and automatically-generated forward inference maps (Fig. 2B) and reverse inference maps (Fig. 2C) for the domains of WM, emotion, and pain. For each pair of maps, similarity was computed by binarizing both maps (i.e., distinguishing between active and inactive voxels) and computing the Pearson correlation (lower triangle) or Jaccard index (upper triangle) across all voxels. Decimals are omitted for legibility. MKDA = multi-level kernel density analysis; FI = forward inference; RI = reverse inference; MA = mega-analysis (for pain, the map reflects pooled estimates from 5 different pain studies rather than an MKDA meta-analysis; see ref[3]). Note the high similarity coefficients for maps within the same domain and low coefficients for pairs of maps from different domains.

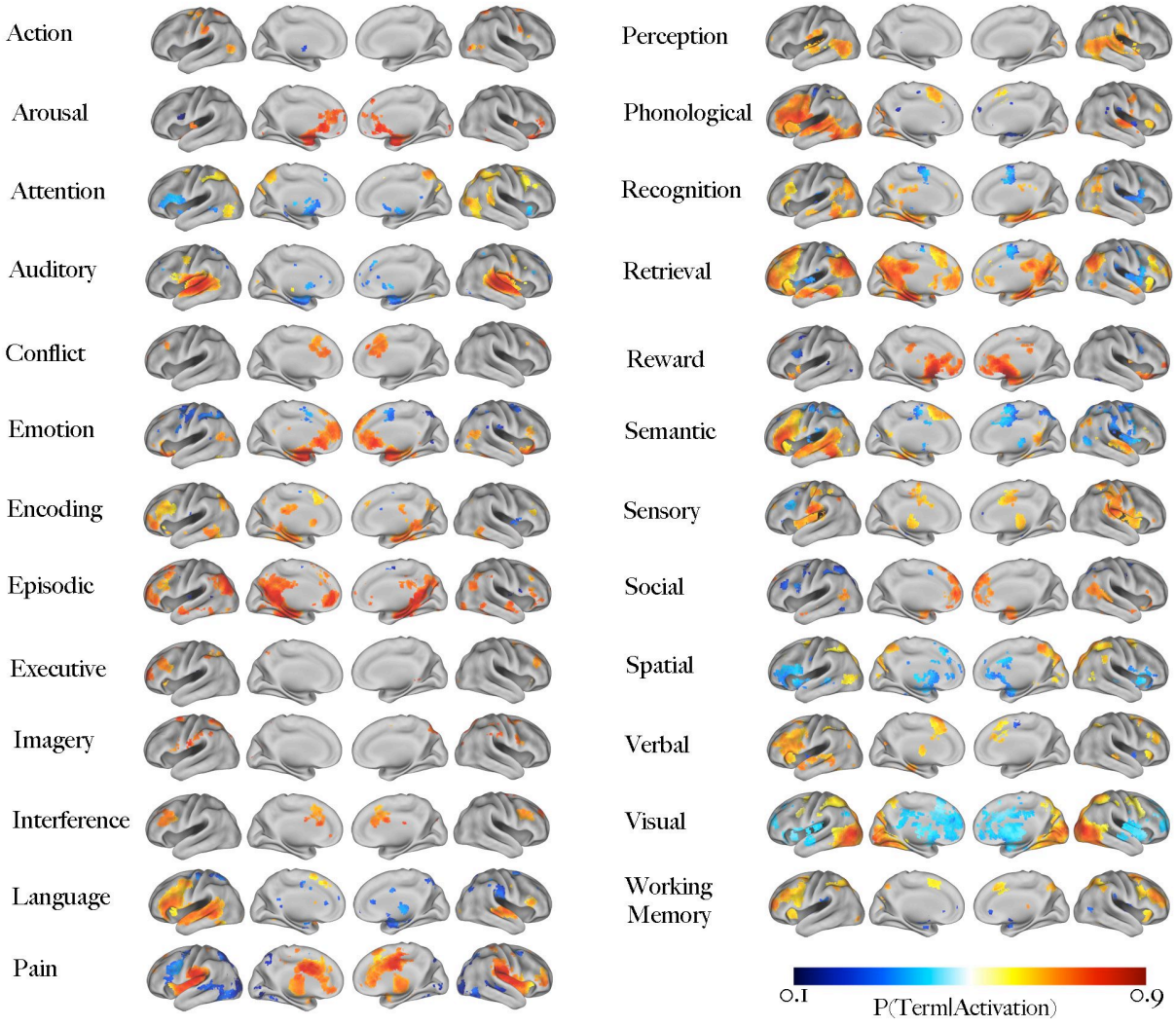
Supplementary Figure 6



Overlap between meta-analyses based on automatically-coded vs. manually-coded pain data. Green: regions associated with the term 'pain' in a fully-automated (forward inference) meta-analysis (slices correspond to surface rendering displayed in Figure 2B). Blue: MKDA[4] meta-analysis results for a manually validated subset of 66 studies drawn from the automatically extracted dataset that contained valid contrasts between pain and a baseline condition. Cyan: overlap of green and blue. To facilitate direct comparison, results of both analyses are thresholded at the same level ($z = 5$). Across voxels, the correlation coefficient (maps unthresholded) and Jaccard similarity index (maps binarized at the $z = 5$ threshold) were 0.84 and 0.65, respectively.

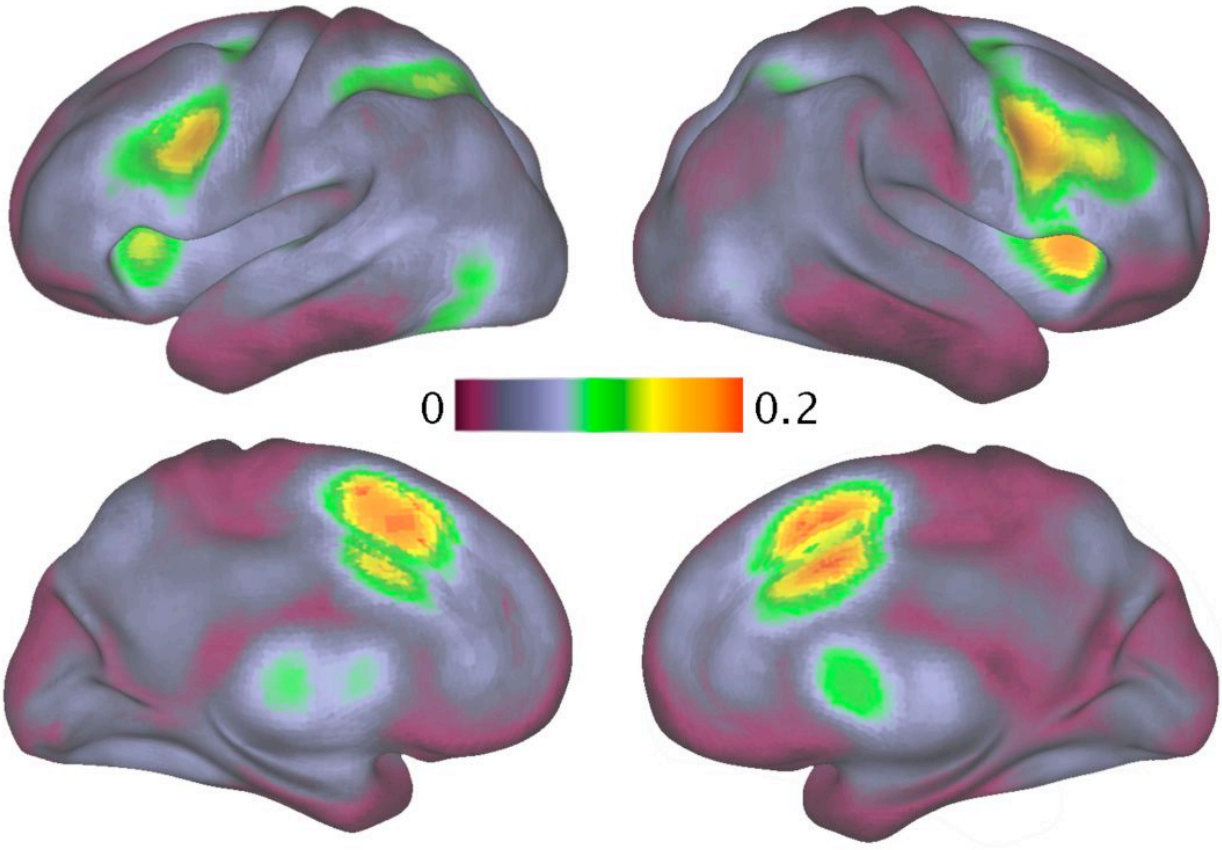
Nature Methods: doi:10.1038/nmeth.1635

Supplementary Figure 7



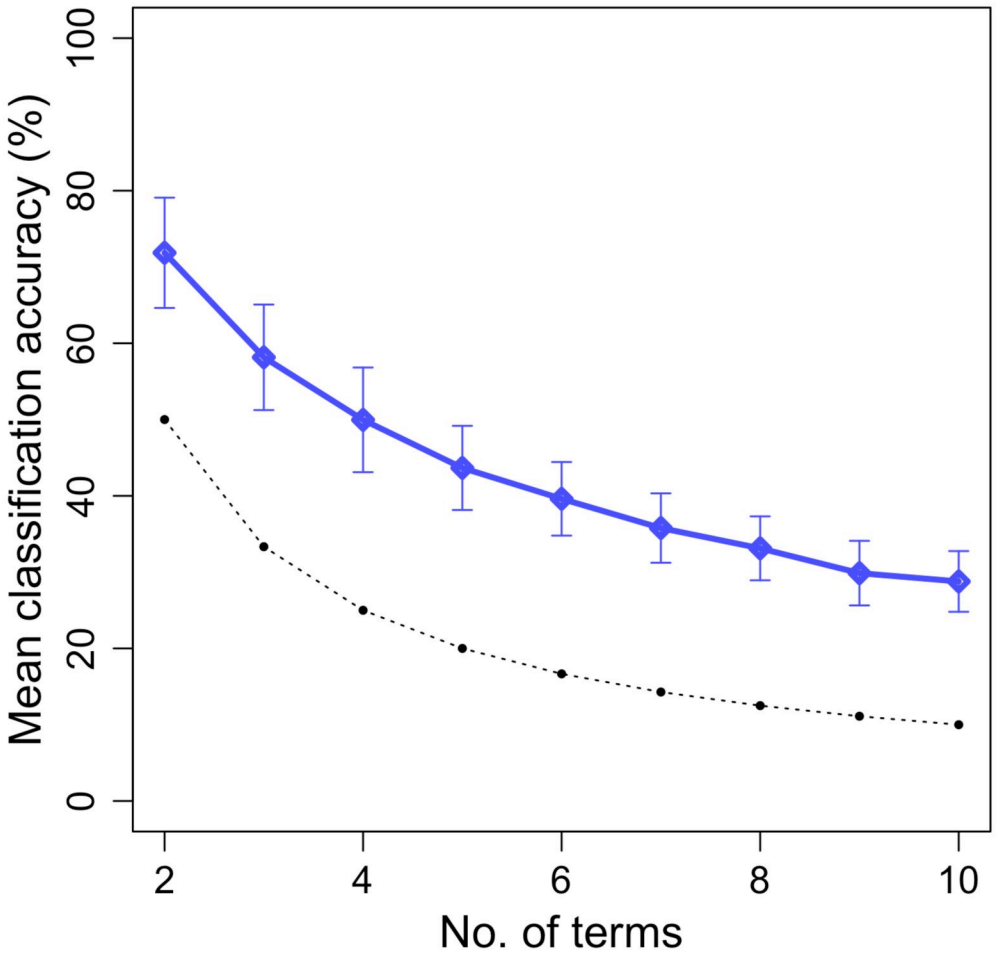
Whole-brain meta-analytic posterior probability maps for 25 key terms that occurred at high frequency (> 1 in 1,000 words) in at least 100 different studies in our database. Voxel values display the probability of the term occurring in a study given observed activation at that voxel (i.e., P(T|A)). To account for base differences in term frequencies, we assume uniform priors for all terms (i.e., equal 50% probabilities of Term and No Term). Activation in orange/red voxels implies a high probability that a term is present, and activation in blue voxels implies a high probability that a term is not present. Values are displayed only for voxels that are significant for a test of association between Term & Activation, with a whole-brain correction for multiple comparisons (FDR = .05). Data available at http://sumsdb.wustl.edu/sums/directory.do?id=8285126.

Supplementary Figure 8



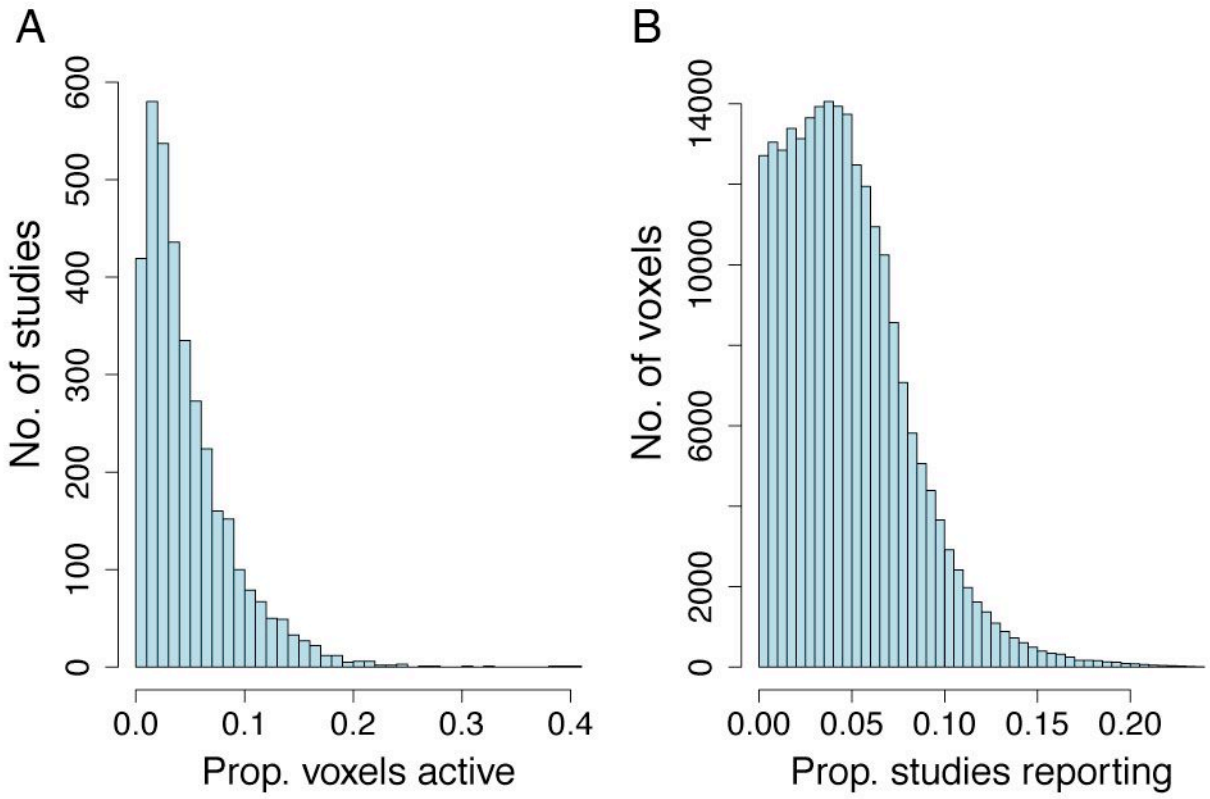
Mean probability of activation at each brain voxel across all 3,489 studies in the database. Frontoparietal regions implicated in cognitive and attentional control were consistently activated at a higher rate than other regions, highlighting the degree to which these areas are nonselectively active (see also Figure 1 in ref.[5]).

Supplementary Figure 9



Accuracy of the naive Bayes classifier as a function of the number of terms being classified. *N*-sized subsets of terms were repeatedly sampled at random from a set of 25 high-frequency terms (cf. Figures 4 and S3). Blue: mean classification accuracy for one hundred random draws. Accuracy was averaged across classes rather than studies to avoid capitalizing on base rate differences between terms. Errors bars reflect the standard deviation. Black: performance level that would be expected by chance.

Supplementary Figure 10



Histograms showing frequency distributions of (A) proportion of voxels reported active in each study and (B) proportion of all studies reporting activity at each voxel. Note that all coordinates reported in articles are convolved with a 10 mm sphere (see Methods); thus, voxels are considered active if they fall within 10 mm of a reported focus.

Supplementary Note

In this note we report additional validation analyses of the NeuroSynth framework. The first section reports a series of supplemental analyses validating the automated coordinate extraction algorithm. The second section reports analyses validating and extending the automated content coding. Throughout both sections, we discuss residual limitations of the NeuroSynth framework and potential directions for future research.

Validation of automated coordinate extraction

As noted in the main text, automated coordinate extraction is susceptible to a number of potential problems that could affect the resulting data quality. First, false positives could occur—that is, the software might incorrectly classify information in a table as an activation focus when it actually represented an entirely different type of information. Second, different software packages and research groups report foci in different stereotactic spaces, resulting in potential discrepancies in the anatomical locations represented by the same set of coordinates across different studies. Third, studies could differ widely in the rigorousness of their experimental and statistical methods, the size of their samples, and the quality of their results, potentially adding noise to the database. Fourth, the software did not discriminate activations from deactivations, and made no attempt to label or categorize foci according to the type of contrast (e.g., task vs. fixation, condition A vs. B, etc.). To address these issues, we conducted a series of additional analyses.

Convergence with manually coded coordinates

To assess the accuracy of the automatic coordinate extraction software, we first compared a set of automatically extracted coordinates with a manually coded set of foci drawn from all studies published in the 2006 and 2007 volumes of Cerebral Cortex and available from SumsDB. Results demonstrated that the automated extraction procedure worked extremely well overall. Eighty-four percent (2929 / 3501) of the foci in the SumsDB reference set were successfully detected by the parser. Inspection of the missing coordinates revealed that in the vast majority of cases, the source of the error was invalid HTML in the table specification. Although we were able to adjust the parser to correctly handle many of these errors, some were idiosyncratic and could only have been handled on a case-by-case basis, which we deemed logistically impractical.

Of the 3334 foci extracted by the automated parser, 405 (12%) were not found in the SumsDB reference set. Careful inspection revealed that 299 of these represented genuine activation foci that were absent from SumsDB; only 106 were 'true' false alarms (though some of these were borderline cases—e.g., foci that were valid, but were from MEG or VBM studies rather than fMRI). Thus, these results suggested an extremely low false positive rate of approximately 3%. This reflects the fact that the parser was designed to be conservative—that is, we deliberately calibrated the software to err on the side of caution

(i.e., to discard any foci that appeared at all questionable). Crucially, there was no reason to expect either false negatives or false positives to produce a systematic bias in our analyses, because the coordinate extraction and semantic tagging procedures were entirely independent of one another. While we are currently working to improve the extraction procedure by developing a more sophisticated parser that uses machine learning techniques to improve its performance with experience, the present results suggest that the current implementation already allows only a small proportion of invalid data into the database.

Contrast-level validation of automatically extracted coordinates

Accurate extraction of coordinates from published articles is necessary but not sufficient for an automated meta-analysis to produce accurate results. If the coordinates extracted by the parser reflect irrelevant experimental contrasts or occur within invalid tables, a meta-analysis could potentially produce null or even misleading results. Because our parser currently lacks the ability to automatically code experimental contrasts, we sought to quantify the loss of signal (if any) associated with the use of a strictly automated approach.

As it was not feasible to manually validate the entire database of over 3,000 studies, we focused on a single psychological domain for which we were able to directly compare automated and manual results. Specifically, we used the NeuroSynth framework to identify 265 studies that used either pain-related terms ('pain', 'painful', 'painfully', 'nociceptive', or 'noxious') or touch-related terms ('touch', 'touched', 'touching', or 'tactile') at high frequency, comprising 8246 activation foci. We then manually inspected and validated all 265 studies. Following inspection, 163 studies (62%) were retained as valid studies that directly investigated pain and/or touch processing (the majority of the 102 excluded studies were relevant to pain or touch—e.g., empathy for pain, real vs. sham acupuncture, etc.—but did not include one or more contrasts directly contrasting relevant pain or touch conditions (e.g., pain vs. rest, touch vs. rest, etc.). Of the 5395 foci automatically extracted from the 162 common studies, 3944 (73%) passed validation and were included in the manually coded dataset (representing 48% of all automatically extracted coordinates).

It is important to note that the majority of the excluded foci were extracted correctly (i.e., the parser identified the correct numerical coordinates), and were excluded because the associated contrast did not meet the stringent criteria of the manual coding. For instance, a large proportion of foci reported pain-related activations at the single-subject (rather than group) level, or reflected contrasts that were only tangentially relevant (e.g., pain empathy vs. rest, heat vs. cold pain, etc.). Although exclusion of these foci from the manual database was clearly the appropriate course, it is important to note that the benefits of an automated approach accrue primarily through large-scale aggregation, and the effects of a decrease in data quality could potentially be offset or even outweighed by the increase in data quantity. That is, excluded studies and coordinates could still contribute useful information to a meta-analysis in the event that the increase in precision and stability of the meta-analytic results provided by larger amounts of data outweighs the increase in noise.

As an empirical test, we compared the (forward inference) pain meta-analysis map produced using the NeuroSynth framework (Figure 2B) with a focused meta-analysis of 66 studies drawn from the manually validated dataset that reported an experimental contrast between painful stimulation and a baseline condition (N = 66 studies reporting pain vs. rest). As Supplementary Figure 6 illustrates, the two approaches produced strikingly similar results (correlation coefficient = .84 across voxels), with partial overlap in virtually all brain regions active in either map. Thus, these results suggest that for broad content terms associated with hundreds of studies and thousands of reported foci, an automated approach can produce largely the same results as a manual approach. However, there is no question that manual coding will continue to be necessary in many if not most cases (e.g., if narrower states such as "pain empathy" are to be distinguished from broad ones such as "pain"), and much of our current work focuses on improving the extraction of metadata that can support more fine-grained automated coding of the experimental contrasts associated with individual activation foci.

Quantification of reported activation increases versus activation decreases

Related to the lack of automated contrast coding is another potential concern that the results generated by our framework inherently confound reported increases and decreases in activation. Because the coordinate parser has no sense of directionality (i.e., it cannot distinguish between task > rest and rest > task), some proportion of the coordinates that reflect reported activation decreases are inevitably treated as activation increases, potentially biasing the results. Because the extent of this problem depends largely on the proportion of total coordinates that constitute activation decreases rather than increases, we sought to quantify the balance between reported increases and decreases in the neuroimaging literature. In lieu of a full manual coding of the entire database, we used the manually-validated pain and touch dataset described above, which included a standardized coding of the contrast corresponding to each valid activation (e.g., pain > rest, high pain > low pain, rest > touch, etc.).

The results of the manual coding indicated that activation decreases were reported much less often than activation increases for all major contrasts. For instance, for comparisons between pain and rest (1863 coordinates), 94% of coordinates were increases (pain > rest) and only 6% were decreases; for comparisons between touch and rest (1293 coordinates), 95% were increases; and for comparisons between high and low pain (548 coordinates), 83% were increases. These findings suggest that, at least for the tested domains of pain and touch, activation decreases constitute a relatively small proportion of activations reported in tables in published neuroimaging articles, and thus appear to exert minimal influence on meta-analytic results (cf. Supplementary Figure 6). Nonetheless, since the extent to which this conclusion holds may vary across psychological domains, future efforts should seek to develop automated ways of coding increases versus decreases or activations versus deactivations—at least for a subset of contrasts that can be relatively easily identified (e.g., those involving terms like 'rest', 'baseline', etc.).

Nature Methods: doi:10.1038/nmeth.1635

Automated stereotactic space detection and coordinate transformation

The Automated Coordinate Extraction software made no attempt to identify the stereotactic space in which coordinates from different studies were reported, or to correct for between-study differences in spaces[6,7]. It is unlikely that such space differences heavily influenced our results, because (a) the great majority of studies (approximately 75 – 80%) are reported using an MNI-based space, and (b) the spatial specificity of meta-analytic results is already limited by the use of a 10 mm smoothing kernel and the marked heterogeneity in preprocessing procedures used in different studies. Nonetheless, the ability to detect and correct for space differences in an automated way would undoubtedly help reduce error by maximizing overlap between coordinates. In an effort to implement an automated space correction procedure, we have begun to develop an algorithm that (a) identifies the stereotactic space used in each study based on the usage of key terms within the article text, and (b) uses an existing affine transformation developed by Lancaster and colleagues[2,7] to convert coordinates from different spaces to a common reference space.

At present, our algorithm distinguishes only between the two most common stereotactic spaces—namely, the MNI-based space used by default in SPM and FSL, and the Talaraich & Tournoux[1] (T88)-based space used by default in AFNI and BrainVoyager. The algorithm assigns the label MNI or T88 to a study in the event that one or more keywords associated predominantly with one space is used at least once in the article text AND there are no occurrences of words predominantly associated with the opposite space. For instance, the occurrence of terms such as 'MNI', 'SPM', and 'FSL' in the absence of any terms like 'Talairach', 'AFNI', or 'BrainVoyager' would be taken to imply that data were reported in MNI space, and vice versa for T88 space (the exception is that the term 'Talairach' is not taken as evidence *for* the use of Talairach space, because many researchers use the term to refer generically to all stereotactic coordinate systems). If the algorithm detects competing evidence (e.g., the terms 'BrainVoyager' and 'MNI' are both used), as might happen when researchers use BrainVoyager software with a non-default template), the label 'UNKNOWN' is assigned (9% of all studies), and no transformation is applied.

To validate the accuracy of our algorithm, we selected a random subset of 100 studies from the database and manually examined each one to identify the originating space. Inspection revealed that the automated space detection algorithm performed relatively well overall. Fifty-eight of 66 (88%) of studies in MNI space were correctly labeled MNI (2 were labeled T88, and 6 unknown), and 15 of 31 (48%) of studies in T88 space were correctly labeled T88 (12 were labeled MNI, and 4 unknown). The relatively high false negative rate for T88 studies was attributable largely to the fact that 10 studies conducted analyses in MNI space but reported transformed coordinates in T88 space. While the results reported here are preliminary, and do not take MNI-to-T88 transformation into account, we anticipate that modifying the algorithm to account for such transformations will be relatively straightforward, as a relatively small set of terms appear to be highly diagnostic (e.g., the terms 'Brett', 'Lancaster', 'converted', or 'transformed' in close proximity to the term 'Talairach'). Thus, we expect a final version of the automated correction algorithm to perform with high accuracy once fully integrated with the NeuroSynth framework. However, it is important to note that our current approach only distinguishes between MNI

and T88 spaces; it does not distinguish between more subtle differences in template (e.g., SPM99, SPM05, and FSL all use slightly different templates by default), and is deliberately conservative, making no attempt to categorize ambiguous studies (9% of studies are labeled unknown). Ongoing work aims to directly address these limitations; in the interim, we expect that relatively limited effort will be required to manually inspect and label the small proportion of studies that use an unclassified space.

Once space labels are assigned, coordinate can be automatically converted between stereotactic spaces using existing affine transformations. Because most studies (approximately two-thirds) report data in MNI space rather than T88, we converted coordinates from T88 studies to MNI rather than the converse so as to minimize error induced by transformation. We used an inverted version of the previously validated icbm_spm2tal affine transformation developed by Lancaster et al[2]. (We chose the SPM version of the transformation rather than the FSL or pooled versions because SPM is by far the most commonly used neuroimaging software package, and the majority of coordinates in published articles are consequently reported for an SPM-based template. Future extensions will support software-specific space detection and transformation.)

To validate the automated application of the Lancaster transformation, we compared the results obtained for studies in TAL space before and after transformation relative to studies in MNI space. Supplementary Figure 1 presents sample results for a reverse inference meta-analysis of the term 'amygdala' with (top) and without (bottom) the transformation. The results demonstrate that differences between MNI and T88 in the spatial localization of the amygdala are substantially reduced following transformation of T88 coordinates, though they remain noticeable, particularly along the rostral/ventral axis. (It is presently unclear whether the residual differences reflect sampling error due to the use of mutually exclusive studies, the misidentification of some T88 studies as MNI studies, or fundamental limitations of the affine transformation, which cannot account for non-linear differences.)

To quantitatively assess the effects of the coordinate transformation algorithm on our meta-analysis results, we automatically generated new meta-analysis maps for 30 common terms drawn from the sets in Figure 5 and Supplementary Figure 2, conducting separate analyses for studies reporting coordinates in MNI space and in T88 space. We then computed the correlation coefficient between the MNI and T88 maps across all voxels both before and after applying the Lancaster transformation. As expected, for virtually all terms (28 of 30), a stronger correlation was observed post-transformation (mean $r = .66$) than pre-transformation (mean $r = .60$; paired t-test, $p < .001$), demonstrating that it is possible to detect and compensate for differences in stereotactic space to a significant extent in an automated way.

Validation and extension of automated content coding

At present, automated coding of article contents is based exclusively on a lexical approach, which assumes that usage rates of individual words provide a reasonable proxy for more effortful manual coding of the psychological processes investigated by individual

neuroimaging studies. The results presented in the main text provide strong support for this assumption, as it would not have been possible to successfully classify study-level and subject-level data if the meta-analysis maps did not accurately reflect stable mappings between cognitive and neural states. Nonetheless, to ensure the accuracy and robustness of the lexical approach we conducted additional validation analyses detailed below.

Convergence with anatomically-defined regions

First, we demonstrated that the lexical approach could recapture conventional boundaries between distinct anatomical regions reasonably accurately. We compared the maps generated using lexical mapping for key anatomical terms (e.g., 'amygdala', 'hippocampus', and 'parahippocampal') with the regional boundaries found in the widely used Harvard-Oxford anatomical atlas[8]. The anatomical labels we used for the lexical meta-analyses were derived from the PubBrain neuroanatomical ontology (pubbrain.org); we selected only those terms that occurred at a high (> 50 studies) frequency, excluding very broad terms (e.g., frontal lobe, telencephalon, etc.). Supplementary Figure 2 displays boundaries for selected regions as defined by the lexical analysis versus the Harvard-Oxford atlas. Because the posterior probability maps could be thresholded arbitrarily, we restricted each lexical ROI to the same number of voxels present in the ROI defined in the Harvard-Oxford atlas. The resulting lexically-defined ROIs were reasonably similar to the corresponding anatomical regions, even for relatively small structures (e.g., the amygdala). Note that dissimilarities between the two maps are not solely attributable to errors in the automatic extraction procedure, as researchers often use anatomical labels somewhat idiosyncratically.

Identification of functionally-selective cortical regions

Second, we used the lexical approach to replicate previous findings of category-specific activation for visual object recognition in posterior cortical regions. The terms 'faces', 'words', and 'places' were strongly and selectively associated with activations in the putative fusiform face area (FFA[9]), visual word form area (VWFA[10]), and parahippocampal place area (PPA[11]), respectively (Supplementary Fig. 3, top). These mappings were further confirmed by searches for 'FFA', 'VWFA', and 'PPA', which aligned closely with the results of the category-based search (Supplementary Fig. 3, bottom).

Convergence with prior literature

Third, we used the lexical approach to generate whole-brain meta-analysis maps for 25 high-frequency terms corresponding to concepts that have been extensively studied in the fMRI literature (Supplementary Fig. 7). The results closely replicated numerous previous studies, with different sets of terms activating expected brain networks. For instance, the terms 'conflict', 'executive', 'interference', and 'working memory' most selectively activated

medial and lateral frontal regions implicated in working memory and executive control; language-related terms such as 'language', 'phonology', 'semantic', and 'verbal' were associated with strongly left-lateralized activations in temporal and ventrolateral prefrontal regions; modality-related terms such as 'visual', 'auditory', and 'sensory' were associated with selective activations in visual, auditory, and sensory cortices, respectively; and so on.

Title-based lexical analyses produce similar results

The empirical evidence reported above suggests that, at least for broad psychological domains, term-based meta-analyses can provide accurate and robust results that converge with prior findings. Nonetheless, it is clear that the present approach has a number of limitations that will be important to address in future work. One is that article coding is currently based solely on the frequency with which terms occurs anywhere in an article and does not take into account contextual information such as the location of a word or its relation to other nearby words. A priori, it is plausible that some article sections (e.g., title, abstract, table captions, or results) might carry more diagnostic information than others (e.g., the introduction or discussion), or that words should be weighted based on their immediate context (e.g., greater weight for words with proximal references to tables or figures). As a preliminary effort in this direction, we conducted a full set of meta-analyses identical to those reported in the text but based solely on the occurrence of words in article titles (rather than anywhere in the text). Despite the large reduction in number of studies associated with each term (most terms, the title-based searches returned only 10 – 20% the number of studies identified by the full-text search), the results were very similar to (if somewhat less sensitive than) those obtained using full-text searches (see Supplementary Fig. 4 for examples), suggesting that the information contained in the full text of articles overlaps closely with that conveyed by article titles.

In future work, we intend to further refine our approach by modeling terms in different article sections separately and empirically identifying weightings that maximize the sensitivity and specificity of the meta-analytic results. Along similar lines, one could also assign different weights to studies—for instance, assigning greater weight to studies that are perceived as more authoritative (e.g., having higher citation counts) or are authored by researchers known to work closely in a particular domain. Our hope is that the availability of the tools, data, and results introduced here will encourage other researchers to contribute to such efforts and implement innovative extensions.

Beyond individual terms

A second limitation of the present implementation is that it is based exclusively on counts of individual words or phrases, whereas the contents of articles are probably better captured using groups of words that naturally coalesce into coherent topics. To facilitate an eventual shift from term-based analyses to topic-based analyses, we have developed a rudimentary syntax for conducting analyses that involve combinations of multiple words.

Our code (available at http://neurosynth.org) embeds a parsing expression grammar[12], enabling users to recursively nest arbitrarily complex queries. For instance, the query:

"(disgust | sad* | anger | fear* | anx*) &~ (pain* | noxious | nocicept*)"

would select all studies that use one or more negative emotion terms (e.g., disgust, sad, sadness, anger, etc.) but NOT one or more pain-related terms (e.g., pain, painful, noxious, etc.). This approach facilitates dynamic analyses that go beyond individual words and allow users to conduct sophisticated meta-analyses targeting arbitrary word combinations. We have also begun to explore other approaches that attempt to model latent topics or clusters in the text of neuroimaging articles (e.g., using latent Dirichlet allocation or multidimensional scaling); however, such analyses are beyond the scope of the present article.

References

1. Talairach, J. & Tournoux, P. *Co-Planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System: An Approach to Cerebral Imaging*. (Thieme: 1988).

2. Lancaster, J.L. et al. Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. *Human brain mapping* **28**, 1194-205 (2007).

3. Atlas, L.Y., Bolger, N., Lindquist, M.A. & Wager, T.D. Brain Mediators of Predictive Cue Effects on Perceived Pain. *Journal of Neuroscience* **30**, 12964 (2010).

4. Wager, T.D., Lindquist, M.A., Nichols, T.E., Kober, H. & Van Snellenberg, J.X. Evaluating the consistency and specificity of neuroimaging data using meta-analysis. *Neuroimage* **45**, 210-221 (2009).

5. Nelson, S.M. et al. Role of the anterior insula in task-level control and focal attention. *Brain Structure and Function* 1-12 (2010).

6. Van Essen, D.C. & Dierker, D.L. Surface-based and probabilistic atlases of primate cerebral cortex. *Neuron* **56**, 209-225 (2007).

7. Laird, A.R. et al. Comparison of the disparity between Talairach and MNI coordinates in functional neuroimaging data: Validation of the Lancaster transform. *Neuroimage* **51**, 677-683 (2010).

8. Smith, S.M. et al. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* **23**, S208-S219 (2004).

9. Kanwisher, N., McDermott, J. & Chun, M.M. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience* **17**, 4302-4311 (1997).

10. McCandliss, B.D., Cohen, L. & Dehaene, S. The visual word form area: expertise for reading in the fusiform gyrus. **7**, 293-299 (2003).

11. Epstein, R. & Kanwisher, N. A cortical representation of the local visual environment. *Nature* **392**, 598-601 (1998).

12. Ford, B. Parsing expression grammars. *ACM SIGPLAN Notices* **39**, 111-122 (2004).