



## Developing a neurally informed ontology of creativity measurement

Yoed N. Kenett<sup>a,1</sup>, David J.M. Kraemer<sup>b,1,\*</sup>, Katherine L. Alfred<sup>b,c</sup>, Griffin A. Colaizzi<sup>d</sup>, Robert A. Cortes<sup>d</sup>, Adam E. Green<sup>d</sup>

<sup>a</sup> Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104, USA

<sup>b</sup> Department of Education, Dartmouth College, Hanover, NH 03755, USA

<sup>c</sup> Department of Psychology, Dartmouth College, Hanover, NH 03755, USA

<sup>d</sup> Department of Psychology, Georgetown University, Washington, DC 20057, USA

### ARTICLE INFO

#### Keywords:

Creativity  
Ontology  
NeuroSynth  
Representational similarity analysis  
RSA  
Multivariate pattern analysis  
MVPA

### ABSTRACT

A central challenge for creativity research—as for all areas of experimental psychology and cognitive neuroscience—is to establish a mapping between constructs and measures (i.e., identifying a set of tasks that best captures a set of creative abilities). A related challenge is to achieve greater consistency in the measures used by different researchers; inconsistent measurement hinders progress toward shared understanding of cognitive and neural components of creativity. New resources for aggregating neuroimaging data, and the emergence of methods for identifying structure in multivariate data, present the potential for new approaches to address these challenges. Identifying meta-analytic structure (i.e., similarity) in neural activity associated with creativity tasks might help identify subsets of these tasks that best reflect the similarity structure of creativity-relevant constructs. Here, we demonstrated initial proof-of-concept for such an approach. To build a model of similarity between creativity-relevant constructs, we first surveyed creativity researchers. Next, we used NeuroSynth meta-analytic software to generate maps of neural activity robustly associated with tasks intended to measure the same set of creativity-relevant constructs. A representational similarity analysis-based approach identified particular constructs—and particular tasks intended to measure those constructs—that positively or negatively impacted the model fit. This approach points the way to identifying optimal sets of tasks to capture elements of creativity (i.e., dimensions of similarity space among creativity constructs), and has long-term potential to meaningfully advance the ontological development of creativity research with the rapid growth of creativity neuroscience. Because it relies on neuroimaging meta-analysis, this approach has more immediate potential to inform longer-established fields for which more extensive sets of neuroimaging data are already available.

Creativity—like all broad psychological constructs—is difficult to pin down (Glăveanu et al., 2019; Glăveanu and Kaufman, 2019; Piffer, 2012; Runco and Jaeger, 2012). A prominent effort to characterize the mental mechanisms that support creative thought, referred to as the *creative cognition* approach, posits that creativity emerges from the interaction of lower level cognitive processes (Abraham, 2013, 2014; Ward, 2007, 1999). Such cognitive processes include memory, reasoning, imagination, and cognitive control (Abraham, 2014; Benedek and Fink, 2019). This approach has been critical for advancing creativity research because it enables researchers to understand the complex, multifaceted nature of this construct by examining its component cognitive mechanisms, which are easier to measure. Additional core psychological abilities have also been identified in relation to creativity, including both divergent and convergent thinking (Guilford, 1950). However, as

the number of cognitive constructs that relate to creativity grows, so does the complexity of operationalizing, measuring, and relating these constructs to each other.

Pinning down creativity by mapping key sub-constructs onto specific experimental tasks is critical to advancing current understanding (Abraham, 2013; Beaty et al., 2019; Benedek and Fink, 2019; Dietrich, 2007, 2019). Rigorous empirical study of creativity, including brain-based inquiry, has recently been recognized as a priority across multiple sectors (Council and Committee, 2005; Florida, 2014; Frey and Osborne, 2017; Jennings, 2010; Lichtenberg et al., 2008; Newcombe, 2017), but lack of clarity concerning measurement in this relatively young field presents a challenge to further progress. Relatedly, there is considerable inconsistency of measurement in the field, such that different groups “research past one another” (Glăveanu et al., 2019) by using different tasks to operationalize the same construct

\* Corresponding author.

E-mail address: [David.j.m.kraemer@dartmouth.edu](mailto:David.j.m.kraemer@dartmouth.edu) (D.J.M. Kraemer).

<sup>1</sup> Both the authors equally contributed to this work.

(Barbot et al., 2019; Cortes et al., 2019). To address this inconsistency, progress toward establishing an ontology of creativity measurement is a priority for advancing creativity research. We use the term, ontology, in the sense of a taxonomic structure that organizes a set of things based on the similarities and differences of their meanings. The meanings of research measures inhere in what their outcomes capture – what they tell us about a person or a group. Thus, an ontology of creativity measurement organizes creativity measures based on the similarities and differences between what each measure captures (i.e., the creativity-related constructs they reflect). Ontologies facilitate coherence within a research field through standardization of constructs and tasks, often requiring large-scale meta-analyses to agree on a set of terms and definitions (Bilder et al., 2009). In the current study, we sought to demonstrate an initial proof-of-concept for a meta-analytic data-driven approach that leverages neuroimaging to support the ontological mapping of creativity-relevant constructs to tasks that measure those constructs.

Ontological development in psychology has historically focused on *a priori* considerations. While considerations of the *a priori* nature and similarity of constructs and sub-constructs are essential to a meaningful ontology, they often prove difficult to objectively weigh against each other, leading to vague or inconclusive outcomes. New resources for aggregating and analyzing neuroimaging data may enable new ways of integrating data-driven approaches with *a priori* considerations toward more objective and more precise ontological development, especially as it regards the mapping of constructs to measurement tools. Cognitive neuroscience has generated a large and growing set of neural data over the course of nearly 30 years, comprising approximately 40,000 studies (Eklund et al., 2016). Substantial research has investigated neural activity associated with a large number of psychological constructs, and an even larger number of specific tasks intended to measure those constructs. Thus, the data now exist to at least begin empirically testing the question, *Which set of tasks reliably elicits neural activity reflective of a given set of cognitive constructs?* In the context of an ontology of creativity measurement, similarity and dissimilarity at the neural level can inform the extent to which different measures reflect similar and/or distinct cognitive constructs. This question is critical for the field of creativity neuroscience research, and psychological research more broadly, and the answers will directly impact our ability to utilize neural data to inform cognitive theories. Tools such as NeuroSynth (Poldrack and Yarkoni, 2016; Yarkoni et al., 2011), a powerful software engine for generating meta-analyses based on text-based searches of thousands of neuroimaging studies, and the BrainMap database (Laird et al., 2005), have been developed in recent years to aid in compiling, analyzing, and interpreting this massive body of data. NeuroSynth allows for comprehensive meta analyses based on selected terms, such as those referring to specific cognitive constructs (e.g. “flexibility”). The resulting meta-analyses indicate areas of the brain that are associated with that particular construct. This outcome is accomplished using brain activation data from all the studies in the database that refer to that particular construct, while controlling for the neural responses associated with every other study in the database (over 14,000 total studies). Researchers can thereby generate new insights about the neural instantiation of specific cognitive constructs, informed by the volume of neural data amassed across thousands of studies conducted over the entire timespan of neuroimaging experimentation.

Here we used this extensive meta-analytic resource to examine a set of constructs that are targets of creativity research. Our main goal was to develop a method by which we can leverage neural meta-analyses to identify a set of commonly-used experimental tasks that elicit neural activity reflective of these cognitive constructs. Our second goal was to examine the ways in which these constructs relate to each other in terms of neural activity—to examine the structure among these constructs on a neural level—in order to inform our understanding of how the brain instantiates creativity as a constellation of constructs. A tertiary goal of our study was to compare this neural construct space to its corresponding construct space defined by creativity researchers, and to examine

similarities and differences in these models in order to learn more about both models and possibly generate hypotheses for further research.

To accomplish these goals, we first generated a model of creativity-relevant cognitive constructs by querying a group of researchers sampled from two academic societies focused on creativity. This model, based on the behavioral ratings of researchers describing the relationship between pairs of constructs, formed our basis for comparison against which we evaluated various models derived from neural data. In order to generate a corresponding neural model, we used NeuroSynth to calculate term-based meta-analyses of neural activity indicating which brain regions are specifically and robustly associated with the same set of cognitive constructs. Inputting these maps into a meta-analytic representational similarity analysis, we then compared the neural data directly to the expert-informed conceptual model. We also generated a separate neural model based on meta-analyses of individual experimental tasks that are commonly used to represent those same constructs. Next, we calculated several variations in both neural models to find a better fit to the expert model. Specifically, we tested whether removing or adding individual tasks or constructs to the neural model improved the fit of the neural data to the expert model. Such an approach allows us to identify the set of tasks best reflecting the similarity structure of the target set of constructs related to creativity.

The focus of this work was on establishing a proof-of-concept for methods that are likely to have long-term value for the selection of experimental tasks to capture given cognitive constructs. Practical constraints, especially concerning the relative paucity of neural data for creativity tasks that are currently available in NeuroSynth, limit the interpretability of the particular set of data we used for this proof-of-concept. However, some preliminary conclusions might usefully be drawn from the results about the constructs and tasks we considered. As the first steps toward data-driven ontological development of creativity research, we explored each of several step-wise methodological approaches briefly in order to demonstrate a proof-of-concept for its use. The main point demonstrated by the current study is that the field of creativity neuroscience is poised to begin a new phase in which a growing volume of available neural data can usefully inform our ontological mappings of constructs to tasks. Using methods such as those described here, we can begin to build an ontology of creativity-relevant cognitive constructs that accurately reflect the brain-behavior relationships described by roughly three decades of empirical observation. These methods can also be used to facilitate similar efforts in other fields.

## 1. Methods

### 1.1. Participants

Sixty-five participants took part in this study. All participants were recruited from academic societies focused on empirical creativity research, the Society for the Neuroscience of Creativity (<https://tsfnc.org>) and the American Psychological Association Division 10 (Society for the Psychology of Aesthetics, Creativity, and the Arts; <http://www.div10.org>). Participants (49% male, 39% female, 12% unknown) had an average age of 39.4 years (SD = 12.21 years) with an average experience of studying creativity of 10 years (SD = 9.5 years). This study was approved by Georgetown University’s Institutional Review Board.

### 1.2. Procedure

Informed consent and all task stimuli were presented via Qualtrics ([www.qualtrics.com](http://www.qualtrics.com)). After providing informed consent, participants were presented with pairs of cognitive constructs from which pairwise ratings were derived, which in turn formed the basis of the expert cognitive model. Participants were first presented with general task instructions indicating the rules of the task and how to record a response for

each item, including an example trial. Following these instructions, participants were presented, one at a time, with each unique pairwise combination of the 10 terms naming cognitive constructs (Cognitive Control, Convergent Thinking, Creativity, Divergent Thinking, Flexibility, Generation, Imagery, Insight, Novelty, Reasoning) for a total of 45 trials. When each pair of terms was presented, participants were also shown a series of seven Venn-diagrams (each containing two overlapping circles representative of the two terms) with varying degrees of overlap, ranging from no overlap to almost complete overlap. The participants were instructed to indicate how much overlap the two terms have by selecting one of the Venn-diagrams by mouse click. Once the participant decided, they were immediately presented with a new pair of terms. Order of term pair presentation was randomized across participants and trials did not advance until participants made a response. Following this task, participants completed other surveys which are not analyzed or discussed further here.

### 1.3. Materials

#### 1.3.1. Expert model of construct space

A behavioral similarity matrix was computed based on participants judging the overlap in similarity of two terms, by selecting from a series of Venn-diagrams (described above), the image that best conveys the similarity of these two terms (Aron et al., 1992; Necka et al., 2015). Participants were presented with all possible pairs of ten terms (45 pairs in total) that named cognitive constructs related to creativity. The ten terms included in this task were: Cognitive Control, Convergent Thinking, Creativity, Divergent Thinking, Flexibility, Generation, Imagery, Insight, Novelty, Reasoning. These terms were part of a large set of possible terms to be used in this task, selected by a group of expert creativity researchers including the authors and the leadership of the Society for the Neuroscience of Creativity. We limited the list of terms to ten terms in order to keep the ratings survey to a manageable length. Furthermore, the ten final terms used in our task were terms that also appeared in the NeuroSynth database, which allowed us to examine how behavioral and neural similarity matrices for the same terms related to each other.

Pairwise ratings of each pair of terms were used to generate a similarity matrix that represents the conceptual space of these 10 constructs. Construction of this type of representational similarity matrix allows for comparison to other data sources, such as neural data, to determine the goodness of fit between two multidimensional representational spaces (Kriegeskorte et al., 2008). In the present study, we use this similarity matrix defined by pairwise expert ratings of creativity-related constructs as a model that represents the way that experts conceive of these constructs in relation to each other. By comparing this model to neural data, as described below, we can evaluate the similarities and differences between the expert conception of these constructs, and the way in which these same constructs manifest in the human brain via the data generated by thousands of neuroimaging experiments.

#### 1.3.2. Term-based meta-analytic maps

In order to create study lists for the ten creativity related terms used in the Creativity Ontology survey, the following steps were taken. In the “initial search” phase, the ten terms used in the survey were entered into NeuroSynth ([www.neurosynth.org](http://www.neurosynth.org); Yarkoni et al., 2011) as a search query for titles in the NeuroSynth database that contained each of the terms. The studies returned by the search were compiled into respective term lists. Next, in the “relevance check cutover” phase, each of the studies were manually checked to ensure that the studies included in the set were appropriate studies of constructs related to creativity (e.g. a study in the Novelty category was about novel uses for items, not about a novel analysis). Term lists with fewer than 20 studies were considered insufficient and removed from further analysis. However, after the cutover, in an effort to gather a sufficient number of studies for Novelty and Divergent Thinking (which initially had fewer than 20 studies each), the terms were queried for titles and abstracts in PubMed in the

“adding PubMed papers” phase. One of the terms, Convergent Thinking, was eliminated from the group for having fewer than 20 studies in total, after exhausting both search methods. After checking all articles (including the PubMed articles) for relevance, the final lists were also checked to ensure that no study appeared in more than one list in the “deleting duplicates” phase. If a study remained on multiple lists, the study was eliminated providing the elimination did not reduce the list below 20 studies. This process resulted in 14 studies remaining on more than one list (but no study remained on more than two lists). Ultimately, the nine remaining terms were: Cognitive Control (63 studies), Creativity (26 studies), Divergent Thinking (20 studies), Flexibility (20 studies), Generation (22 studies), Imagery (44 studies), Insight (21 studies), Novelty (20 studies), Reasoning (46 studies). See SI Table 1 for the full list of PMIDs included for each term, and SI Table 2 for a sample of papers included for each term.

All NeuroSynth based analyses were run on a local implementation of the NeuroSynth core tools (<https://github.com/neurosynth/neurosynth>) using the database version 0.7, released in July 2018 and includes activation data from 14,371 studies. All neuroimaging data and images from NeuroSynth are previously registered to 2 mm MNI space. Because the NeuroSynth database includes data from some non-fMRI neuroimaging studies (e.g., PET, or voxel-based morphometry; for details on data selection see Yarkoni et al., 2011), our selection criteria resulted in a small number of these non-fMRI studies being included in our analyses. These studies represent a small percentage of the total number of included studies (e.g., only 4 studies out of approximately 500 contained PET data), and are noted in SI Table 1 and SI Table 2. The numbers of these studies for each non-fMRI source of data were too small to reliably determine whether the signal provided by these sources differed significantly from the fMRI data. However, in general, inclusion of multiple converging sources of data should increase the power to detect meaningful signal related to cognition-brain associations.

For each of the nine terms, a term-based meta-analysis was conducted using the activation data associated with the PubMed IDs for each of the studies (see Fig. 1 for examples). Both association and uniformity test z-maps were generated for each term-based meta-analysis, and each z-map is FDR corrected at  $p < .01$ . Whereas uniformity tests indicate which brain-regions show consistency of activation within the set of included studies, the association test shows which brain regions have higher levels of activation in the set of included studies compared to the rest of the studies in the full database (Yarkoni et al., 2011). Therefore, the association FDR-corrected z-map was chosen for further analysis.

#### 1.3.3. Neurally-defined model of construct space

In order to compare the expert-generated behavioral ratings described above to a similar model of construct space defined by neural data, we generated a neural similarity matrix (Fig. 2) as follows. For each of the 9 term-based meta-analytic maps described in the previous section, the brain maps were converted into vector arrays, where each element in the array represented the FDR-corrected z-value for the meta-analytic map at that voxel. Then we calculated the Spearman correlation between that neural map and every other meta-analytic neural map. In this way, we were able to obtain a single value representing the Spearman correlation between every pair of meta-analytic maps (36 unique correlations in total). These correlation values were then input into a representational similarity matrix, in which each cell represents the correlation between two meta-analytic maps, and the entire matrix thus comprises every possible pairing of meta-analytic maps (Fig. 2).

#### 1.3.4. Task-based meta-analytic maps

Meta-analytic task-based maps were created through NeuroSynth, similar to the term-Based maps described in the preceding section. For each of the creativity-related constructs included in the expert model, a list of tasks commonly used to measure each construct was generated by the authors (Table 1). Each author listed tasks that they believe best measure the construct terms, based on use in previous literature. Tasks

**Table 1**

Candidate tasks commonly used to operationalize each cognitive construct.

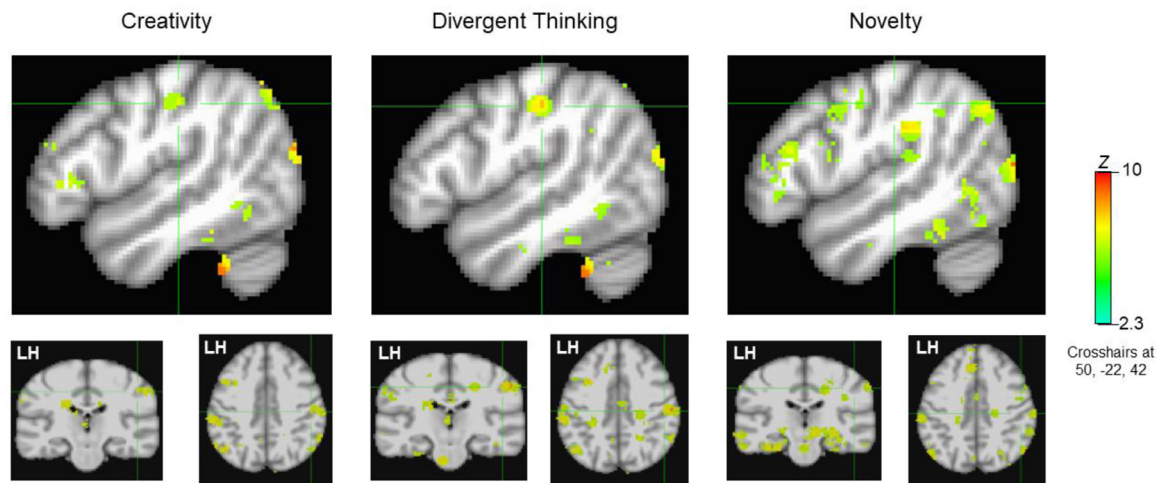
Cognitive Control	Creativity	Divergent Thinking	Flexibility	Generation	Imagery	Insight	Novelty	Reasoning
Stroop	Analogical Reasoning	Verb Generation	Task Switching	Fluency	Mental Rotation	Analogical Reasoning	Verb Generation	Analogical Reasoning
Flanker	Verb Generation	Fluency	Wisconsin Card Sorting	Verb Generation		Wisconsin Card Sorting	Fluency	Wisconsin Card Sorting
Go/No-Go								

**Table 2**

Spearman correlations between expert ratings and neural data for similarity matrices and individual terms.

	Entire Matrix	CC	Cr	DT	Fl	Gn	Im	In	Nv	Rs
Full Model	.21	<b>.67*</b>	.44	<b>.89**</b>	.03	.38	−0.25	.05	<b>.72**</b>	.39
<i>Leave-one-out models:</i>										
Cognitive Control	.09	NA	.55	<b>.86**</b>	.12	.48	.17	.02	.60	.13
Creativity	.30	<b>.68*</b>	NA	<b>.92***</b>	.19	.55	−0.15	.31	<b>.61*</b>	.59
Divergent Thinking	−0.12	<b>.71*</b>	.37	NA	−0.12	.19	−0.18	−0.01	<b>.62*</b>	.34
Flexibility	<b>.35*</b>	<b>.68*</b>	<b>.66*</b>	<b>.88**</b>	NA	.55	−0.19	.10	<b>.81**</b>	.45
Generation	.23	<b>.71*</b>	.46	<b>.93***</b>	.10	NA	−0.24	.03	<b>.76*</b>	.45
Imagery	<b>.38*</b>	<b>.86**</b>	.53	<b>.85**</b>	.00	.38	NA	.08	<b>.79*</b>	.54
Insight	.32	<b>.63*</b>	.47	<b>.90**</b>	.21	.31	−0.17	NA	<b>.81**</b>	.49
Novelty	.06	<b>.64*</b>	.25	<b>.88**</b>	.10	.30	−0.18	.23	NA	.37
Reasoning	.23	.53	.45	<b>.86**</b>	.07	.43	−0.19	.06	<b>.79*</b>	NA

Note: \* =  $p < .05$ ; \*\* =  $p < .01$ ; \*\*\* =  $p < .001$ . All  $p$ -values are generated by permutation correction against a null distribution of 10,000 random permutations. For the Full Model, all constructs were included in both the expert and neural similarity matrices. Row labels for the leave-one-out models indicate the term left out of that model. Column labels indicate the term of interest being correlated between neural and behavioral data sources, each drawn from the same model. CC = Cognitive Control; Cr = Creativity; DT = Divergent Thinking; Fl = Flexibility; Gn = Generation; Im = Imagery; In = Insight; Nv = Novelty; Rs = Reasoning.

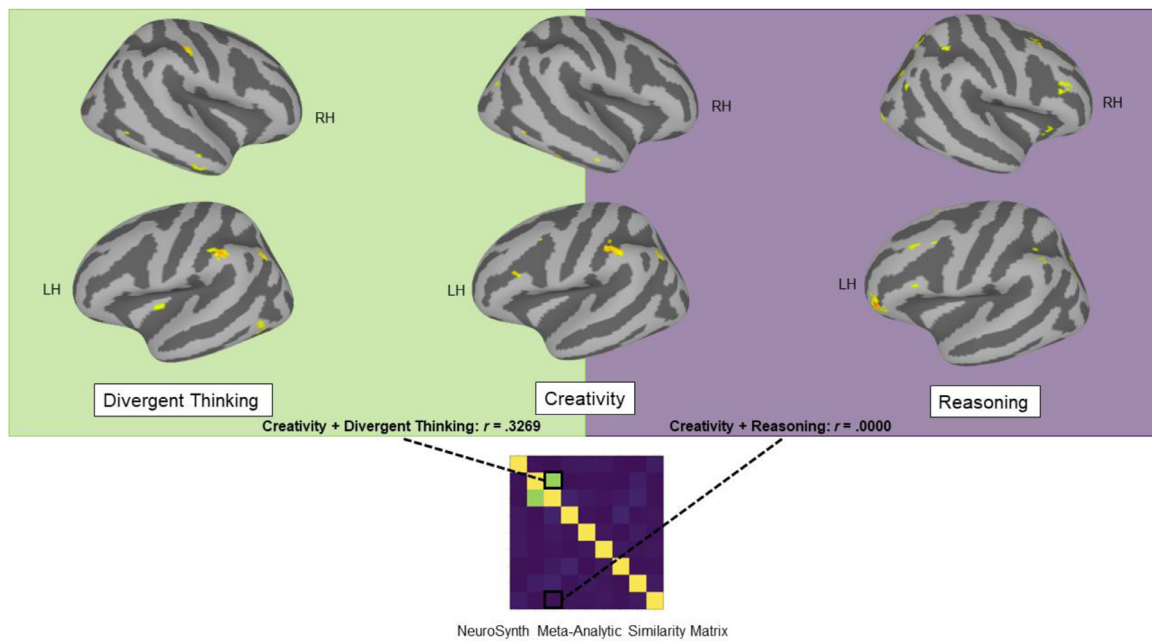


**Fig. 1.** Term-based meta-analytic association Z-maps for Creativity, Divergent Thinking, and Novelty. Uncorrected Z maps generated using NeuroSynth are shown here to display full results; FDR-corrected Z-maps are displayed in Fig. 2 and were used for all analyses. Each map is thresholded at  $Z > 2.3$  and spatially clustered in the volume resulting in a minimum cluster size of 20 voxels per cluster.

were allowed in multiple lists as long as the task is commonly used to measure all of the constructs it is listed for, but any given study could only be present in one task list. We then conducted a PubMed search for each task along with the term “fMRI” to identify neuroimaging studies using these tasks in the “Initial PubMed Search” phase. A list for each of the tasks was generated using the PubMed IDs for each of the studies that met that search query. These studies were initially culled to only include studies that appeared in the NeuroSynth database (limited to studies that have been processed to catalog the neuroimaging regions for results) in the “Cross-reference with Neurosynth” phase before being manually reviewed to ensure the study was appropriately related to creativity and the task was used for the neuroimaging results in the

“Relevance Check Cutdown” phase. Finally, any duplicate studies were removed from all instances on the task lists and the two list sets (task and term) were compared to identify any duplicate studies between the sets in the “Deleting Duplicates” phase. If a study existed across both sets, the study was eliminated providing the elimination did not reduce the term list below 20 studies or the task list below 10 studies. This process resulted in no duplicate studies between task lists, and 26 duplicate studies remaining across sets (i.e., between task lists and construct lists). See SI Table 3 for the full list of PMIDs included for each task, and SI Table 4 for a sample of papers included for each task. With those lists of IDs, association and uniformity test meta-analytic z-maps were generated for each of the tasks, and were FDR corrected at  $p < .01$ . As noted





**Fig. 2.** Meta-analytic maps for some terms showed more pattern similarities than others. The NeuroSynth meta-analytic maps for Creativity and Divergent thinking show some regions of overlap and have a correlation of  $r = 0.33$ . By contrast, Creativity and Reasoning have no overlap in patterns of activity, and do not correlate. All meta-analytic surface-based Z-maps were generated in 3-dimensional MNI space using the NeuroSynth association map function and FDR-corrected ( $p < .01$ ).

**Table 3**

Spearman correlations for the models using individual neuroimaging tasks to represent cognitive control.

	Flanker	Go/No-Go	Stroop
Full Model	.01	.10	.13
Individual constructs:			
Cognitive Control	.24	.00	<b>.63*</b>
Creativity	<b>.66*</b>	<b>.66*</b>	.33
Divergent Thinking	<b>.92**</b>	<b>.93**</b>	<b>.83**</b>
Flexibility	−0.03	.03	.08
Generation	.42	.38	.28
Imagery	−0.25	−0.23	−0.09
Insight	−0.18	−0.02	−0.02
Novelty	.53	<b>.65*</b>	<b>.65*</b>
Reasoning	.34	−0.08	.39

Note: \* =  $p < .05$ ; \*\* =  $p < .01$ . All  $p$ -values are generated by permutation correction against a null distribution of 10,000 random permutations. Row labels indicate the terms correlated between behavioral and neural data sources. Column labels indicate the task used to represent *cognitive control* in each neural model.

in the previous section, the FDR corrected association map for each of the tasks was used for further analysis.

#### 1.4. Statistical analyses

##### 1.4.1. Multidimensional scaling analysis

For the purpose of illustrating the relationships between terms as defined by the expert model and the neural model, we conducted a classical (metric) multidimensional scaling analysis (using the *cmdscale* function in R from the *stats* package; <https://www.rdocumentation.org/packages/stats>). This analysis, used mainly to depict similarities and differences between the models, generated a projection of each construct into 2-dimensional space, based on the similarity matrices described above. The data points in this 2-dimensional space were then subjected to a  $k$ -means clustering algorithm (using the *kmeans* function in R from

the *stats* package; <https://www.rdocumentation.org/packages/stats>) aimed at defining up to 3 distinct clusters.

##### 1.4.2. Spearman correlations and permutation corrections

We used Spearman correlation to test the fit between the similarity matrix generated by behavioral ratings of experts and the similarity matrix generated by neural meta-analyses. Spearman correlations were also used to test the fit between the modified similarity matrices described below. All correlations were permutation-corrected to determine significance. When each correlation was calculated, we randomized one of the two matrices or vectors (depending on the analysis) 100,000 times and re-ran the correlation to generate a distribution of potential correlations values from the distribution of our data. From that distribution, we z-scored the actual observed correlation value to identify where it fell relative to the distribution of permuted correlation values, and therefore how likely it was that we found the observed correlation by chance, given our data. This approach allows us to correct for multiple comparisons without making any assumptions about the distribution of our data.

##### 1.4.3. Row-wise analysis

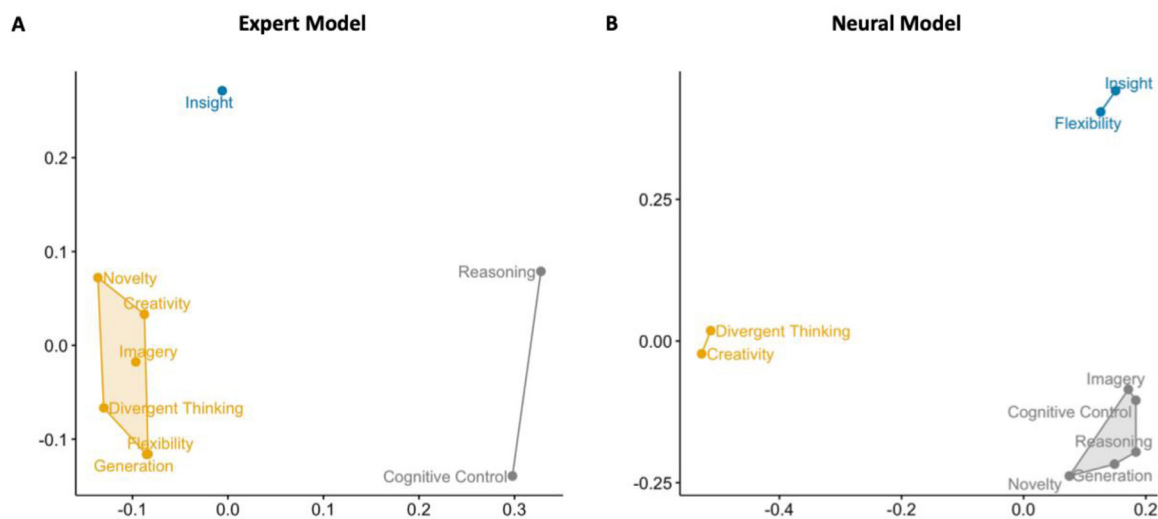
When correlating between full models, we make use of the full representational similarity space between behavioral ratings and patterns of neural activity. However, this analysis provides only one correlation value to represent all of the data contained in both of those similarity matrices, including all 9 constructs. To examine specifically how similar *creativity* is, for example, when comparing its place in the behavioral construct space to its place in the neural construct space, we use row-by-row correlations (i.e., individual term correlations). In these analyses, we correlate the vector associated with each term drawn from the behavioral similarity matrix with the same vector drawn from the neural similarity matrix. This approach allows us to identify the correlation between the behavioral and neural data sources for each construct and for each task.

**Table 4**

Full model and term-wise Spearman correlations for the models using tasks to represent all constructs.

Cognitive Control Tasks	Flanker	Verb Gen	Go/No-Go	Verb Gen	Stroop	Verb Gen
Creativity Tasks	Analogy		Analogy		Analogy	
Full Model	−0.15	−0.25	−0.37	<b>−0.52*</b>	−0.24	−0.32
<i>Behavioral Ratings by Constructs:</i>						
Cognitive Control (varies by model)	.60	.71	.20	.71	.37	.37
Creativity (varies by model)	.46	−0.12	.46	−0.12	.46	−0.12
Flexibility (Task Switching)	.37	.09	.37	.09	.37	.26
Generation (Fluency)	.43	.54	.20	.54	.20	.31
Imagery (Mental Rotation)	−0.12	−0.03	−0.06	−0.03	.06	.32
Reasoning (Wisconsin)	.49	.70	.12	.70	.12	.32

Note: \* =  $p < .05$ . All  $p$ -values are generated by permutation correction against a null distribution of 10,000 random permutations. Table headings refer to the model used to generate the NeuroSynth Neural Task similarity data with the construct-specific behavioral ratings of similarity. Cognitive control was represented by each of: Flanker, Go/No-Go, and Stroop tasks. Creativity was represented by each of: Analogical Reasoning and Verb Generation tasks. All other constructs were represented by an individual task each, indicated in parentheses after the construct name. Under each model heading, each cell contains the Spearman correlation between the Neural Task similarity for that model and the behavioral rating similarity. For example, the cell for Reasoning under Stroop/Analogy indicates the correlation between the Wisconsin Card Sorting task and Reasoning ratings while Cognitive Control is represented by the Stroop Task and Creativity is represented by Analogical Reasoning. Analogy=Analogical Reasoning; Verb Gen=Verb Generation.



**Fig. 3.** Multidimensional scaling plots demonstrating the structure of construct space in two principal dimensions. A: Projection of expert model reflecting construct space defined by pairwise similarity ratings of creativity researchers; B: Projection of neural model reflecting construct space defined by pairwise similarity of NeuroSynth meta-analytic association maps generated by term-based meta-analysis. Colors depict results of k-means clustering.

#### 1.4.4. Leave-one-out analysis

To examine the contributions of each of the terms to the overall model fit between behavioral and neural data sources, we conducted a leave-one-out analysis. We conducted this analysis by iterating through the model and leaving out one term in each iteration and then calculating a new Spearman correlation between the revised NeuroSynth-defined similarity matrix (i.e., the neural leave-one-out model) and the revised behavioral similarity matrix (i.e., the expert leave-one-out model). A total of nine additional models were generated thusly.

## 2. Results

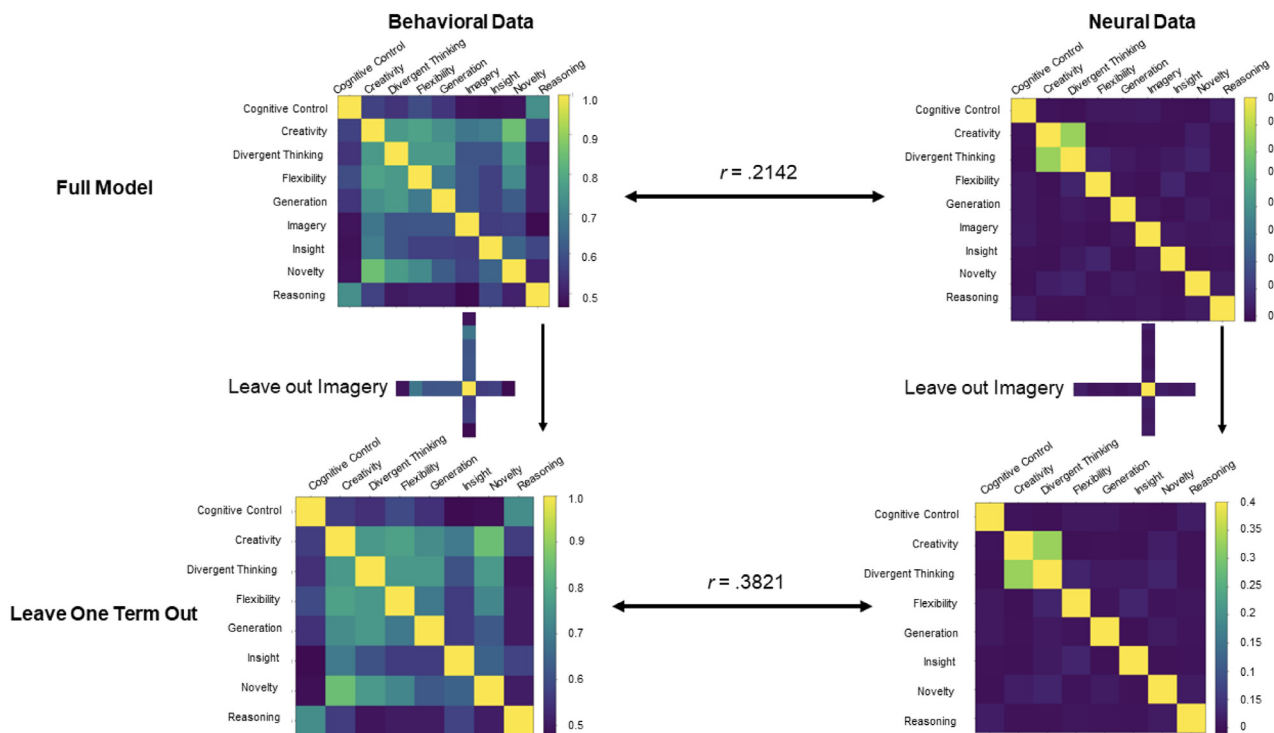
### 2.1. Relating the meta-analytic neural model to the expert-based construct similarity model

#### 2.1.1. Full model space

As described above, our primary goal was to develop a means of using neural data to inform our understanding and measurement of the cognitive constructs that comprise creativity. We began with a quantification of the similarity ratings of experts regarding the relationships between these constructs, thus defining a multidimensional space of cog-

nitive constructs (Fig. 3). As our first approach to testing this expert-defined construct space against neural data, we used NeuroSynth to generate term-based meta analyses for every construct in the expert model test (see Methods section for details). Each construct was used as the basis for a separate whole-brain meta-analysis, and the results were then combined into a full model indicating the neurally-defined construct space (Fig. 3). We then compared this neural model to the expert-defined cognitive model using Spearman correlation (Fig. 4). The correlation between these similarity matrices is the most direct test of whether the data generated by the field of cognitive neuroscience reflects the way that experts conceptualize the space of creativity-related constructs. Results revealed a non-significant correlation between the full expert model and the full neural model,  $r(34) = 0.21$ ,  $z = 1.26$ ,  $p = .10$ .

To gain further insight into the mapping between expert and neural models, we conducted a multidimensional scaling analysis to illustrate the construct spaces defined by each model. As seen in Fig. 3, some constructs anchor the multidimensional construct space similarly in both the neural model and the cognitive model. For example, in both models, the term *creativity* clusters with the term *divergent thinking*. Likewise, in both models, the terms *reasoning* and *cognitive control* cohere



**Fig. 4.** Behavioral and neural results for the full model space and a reduced set of constructs. Top row, left: Similarity matrix defined by experts through pairwise ratings of terms. These ratings are scaled to range from 0 to 1 (1 = complete conceptual overlap between terms). Top row, right: Similarity matrix defined by NeuroSynth term-based meta-analyses. Each of the FDR-corrected NeuroSynth meta-analytic association z-maps were Spearman-correlated with each other to create a measure of how similar the patterns of neural activity associated with those terms are to each other. Bottom row, left: A reduced construct space is defined by removing the term imagery from the expert similarity matrix. Bottom row, right: A reduced construct space is defined by removing the term imagery from the neural similarity matrix. In this example, correlation between the expert model and neural model improves when the construct imagery is removed. Note: Each similarity matrix is scaled separately to better illustrate variations in patterns of similarity. Full similarity matrices are shown here, however values along the diagonal were excluded from analyses.

into the same cluster. Also, in both models, *insight* falls into a separate cluster from either of these two other clusters, falling somewhere in between these other terms along one dimension of the 2-dimensional projection space. However, many differences are also notable between the two models (Fig. 3): *Critically, we believe that the differences between the expert model and the neural data are at least as informative as the similarities between them.* Next, we examine the correlation between each individual construct as defined by these two separate multidimensional spaces, as well as the effect that each construct has on the overall fit between the two models.

## 2.2. Individual term correlations and leave-one-out models

To the degree that there was not a perfect correlation between the neural model and the expert model, there are many potential sources to explain this disconnect. As evident in Fig. 3, several cognitive constructs were not well represented by the neural data. The misalignment of even one construct substantially reduces the goodness-of-fit for the overall model, and this reduction is compounded by the aggregate of several misaligned constructs. Therefore, in the next step of this analysis, we examined the correlation between individual terms as defined separately by the neural model and the cognitive model. To further investigate the goodness of fit between expert-generated cognitive constructs and their neural counterparts, we examined how well each individual construct as defined by the expert model correlated with the same construct as defined by the neural model. Along these same lines, we also tested whether the overall model fit between the expert model and the neural model improved when removing each individual construct from the model (Fig. 4). This approach has the long-range potential to reveal im-

portant information to the field; namely, which constructs are not isomorphic between the conception of expert researchers and the observed patterns of neural activity from the aggregated results of the field as a whole.

To examine the overall improvement in fit between the models generated by the expert ratings and by the neural data upon removing each individual construct, we iterated through each construct in succession removing it from both models (expert and neural) and then correlating the resulting expert and neural similarity matrices leaving out only that one construct at a time. The results of these Spearman correlations are reported in the left column of Table 2. Results demonstrate that for two terms, *flexibility* and *imagery*, removing either term from both the expert similarity matrix and the neural similarity matrix increases the correlation strength between these two multidimensional spaces. When *flexibility* is left out of the model, the expert ratings and neural data correlate at  $r(28) = 0.35$ ,  $z = 1.87$ ,  $p = .03$ . Likewise, leaving *imagery* out of the model results in a significant correlation between expert and neural data  $r(28) = 0.38$ ,  $z = 1.95$ ,  $p = .03$ . These results demonstrate that when these individual constructs were included in the full models, they each contributed to reducing the overall goodness of fit between the two representations of the construct space; i.e., both constructs decreased the correlation between the model generated by the expert ratings and the model generated by the neural data. More generally, these results also confirm that this approach can be used to detect changes in an overall representational space that result from removing an individual construct of interest.

Next, to examine the relationship of each construct in the expert model to each of the constructs in the neural model, we computed correlations between the rows of each similarity matrix using the full expert

model and the full neural model (see Table 2, top row). In this way, each term was defined as a point in the construct space defined by each data source (expert ratings and neural data), and the similarity between the vector coordinates of each of these points was calculated using Spearman correlation. This analysis revealed the strongest neural-behavioral correlations for the terms Cognitive Control,  $r(8) = 0.67$ ,  $z = 1.91$ ,  $p = .03$ , Divergent Thinking,  $r(8) = 0.89$ ,  $z = 2.53$ ,  $p = .01$ , and Novelty,  $r(8) = 0.72$ ,  $z = 2.05$ ,  $p = .02$ . Weaker correlations were observed for the terms Creativity,  $r(8) = 0.44$ ,  $z = 1.27$ ,  $p = .10$ , Generation,  $r(8) = 0.38$ ,  $z = 1.09$ ,  $p = .14$ , and Reasoning,  $r(8) = 0.39$ ,  $z = 1.12$ ,  $p = .13$ . Correlations in the null to negative range were observed with the terms Flexibility,  $r(8) = 0.03$ ,  $z = 0.09$ ,  $p = .46$ , Insight,  $r(8) = 0.05$ ,  $z = 0.14$ ,  $p = .44$ , and Imagery,  $r(8) = -0.25$ ,  $z = -0.72$ ,  $p = .24$ . This analysis demonstrates both convergence and divergence between expert and neural models on the level of individual terms, which is a useful demonstration for future uses of this methodology. However, given the small number of terms in the present models—and consequently few degrees of freedom in the present analyses—the statistical significance of these correlations should be viewed with caution. Table 2 also shows the results for each term from each of the other leave-one-out models, but for space considerations (and because these results do not directly relate to the goals of the current research) these results are not interpreted in further detail here.

### 2.3. How much depends on the selection of a task to represent a cognitive construct?

To define our initial neural model, the analysis above relied on meta-analyses based on terms that named cognitive constructs. It is a strength of tools such as NeuroSynth that we can now easily generate such term-based meta-analyses on the level of whole constructs. However, in any given study, each construct of interest is typically operationalized through a single specific experimental task. Clearly, the choice of task used to represent a given construct has a fundamental effect on the resulting neural activity, and there can be great variability between two tasks that claim to measure the same cognitive mechanism (Poldrack et al., 2011). Therefore, for our second analysis approach, we chose to go beyond entire constructs and focus on individual tasks by testing the change in correlation between the expert model and neural model when a single task is taken to represent a given construct. This approach takes the neural model created in the first analysis and replaces one of the term-based meta-analysis (e.g., for the construct, *cognitive control*) with a task-based meta-analysis focusing on a single task that is often used to represent that construct (e.g., “Stroop color-word”). In this way, we can estimate how much the correlation between the expert model and the neural model improves or worsens as we replace a construct with a specific task, and as we replace each individual task with another task (e.g., replacing “Stroop” with “Go-NoGo” or “Flanker”), while holding the rest of the representational space constant. Therefore, the results of this analysis demonstrate the effect of choosing a single task over another task as a stand-in for an entire construct, with respect to other constructs related to creativity.

As a demonstration of this method, we chose to use three common tasks used to operationalize *cognitive control*: the Flanker task (Chen et al., 2015; Grajewski et al., 2011; Wager et al., 2005), Go/No-Go task (McCormick et al., 2016), and Stroop (Liu et al., 2015; Shin and Kim, 2015). We then correlated the revised neural similarity model (now with 8 term-based neural meta-analytic maps and one task-based meta-analytic map) with the similarity matrix made from the experts’ behavioral similarity ratings. This analysis allows us to consider whether removing unnecessary noise from using a variety of different tasks to measure the same construct results in a better fit with the expert model. Alternatively, the meta-analytic neural maps used in the prior analysis (which include a variety of tasks) could result in a more robust neural signal that better fits the expert model. Finally, as in the analysis above,

we compute the row-wise correlation for each of the terms and tasks (Table 3).

The Spearman correlation analysis of the full model replicated the previous results, in the sense that none of the correlations were significant (based on the permutation test, all  $p$ 's > 0.1). In fact, the effect sizes of each of these correlations is nominally lower than the original model correlation above. For the row-wise correlation analysis, however, we found both increased and decreased neural-behavioral correlations across the three models. To answer the question of which task best represents the construct of cognitive control, we only find a significant neural-behavior correlation for *cognitive control* in the Stroop model,  $r(8) = 0.62$ ,  $z = 1.79$ ,  $p = .04$ , and not in the Flanker model,  $r(8) = 0.24$ ,  $z = 0.69$ ,  $p = .24$ , or the Go/No-Go model,  $r(8) = 0.00$ ,  $z = 0.001$ ,  $p = .99$ . This result suggests that among these three tasks, using the Stroop task to represent *cognitive control* represents the best alignment between neural and behavioral models; i.e., this task may elicit a neural response that most closely reflects the way in which creativity experts conceive of cognitive control. However, we only find significant neural-behavior correlations for the construct *creativity* in the Flanker,  $r(8) = 0.66$ ,  $z = 1.89$ ,  $p = .03$ , and Go/No-Go models,  $r(8) = 0.66$ ,  $z = 1.89$ ,  $p = .03$ , and not in the Stroop model,  $r(8) = 0.33$ ,  $z = 0.95$ ,  $p = .17$ . Thus, the model that produces the best fit for *cognitive control* produces the worst fit for *creativity*.

Another notable result is that for all three models the neural-behavior row-wise correlation of *divergent thinking* were significant (Flanker:  $r(8) = 0.92$ ,  $z = 2.63$ ,  $p = .004$ ; Go/No-Go:  $r(8) = 0.93$ ,  $z = 2.65$ ,  $p = .004$ ; Stroop:  $r(8) = 0.83$ ,  $z = 2.34$ ,  $p = .009$ ). However, we only found significant neural-behavior correlations for *novelty* for the Go/No-Go,  $r(8) = 0.65$ ,  $z = 1.86$ ,  $p = .03$ , and the Stroop,  $r(8) = 0.63$ ,  $z = 1.86$ ,  $p = .03$ , models. Overall, this analysis demonstrates that whereas all three of the selected experimental tasks are commonly used to measure *cognitive control*, each of the tasks captures somewhat different neural-behavior relationships.

We further investigated the impact of task selection on fit with the expert-defined model by conducting a third analysis in which we replaced every construct with a task that is commonly used to represent that construct. Whereas the second analysis method (above) used mostly construct-based neural meta-analyses into which we slotted one task-based neural meta-analysis to stand in for a corresponding construct, this third approach uses only task-based meta analyses to define the neural model. In other words, for each construct in the expert model, we generated a neural meta-analysis based on a single task to represent the neural counterpart of the corresponding cognitive construct. There are two main motivations for this approach. One motivation is that most neuroimaging studies use only a single task to probe a cognitive construct, so defining the neural model exclusively by the most-often-used single tasks provides a means of demonstrating how well any individual study might approximate the expert construct space. This approach therefore serves as a useful complement to the analyses above.

A second—and related—motivation is that perhaps when our expert participants generated their ratings for each construct, they were basing these ratings (at least in part) on a set of commonly-used tasks. If true, this interpretation raises the possibility that a neural model defined by using only one task to represent each construct might be a better match for the expert model. In fact, even disregarding the various ways in which experts may have interpreted the task instructions, it is possible that using a single task to define each construct reduces the variance due to noise that results from incorporating so many different tasks into the meta-analytic model. In short, it may be that the neural space of these constructs is better defined by a more constrained set of tasks that powerfully tap the targeted cognitive mechanisms.

To test this possibility, for our third analysis approach we examined how well the tasks that are used to represent the terms capture neural-behavior relations. To do so, we identified up to three common tasks for each of our nine terms (Table 1). This process led to three of the terms to be rated as having the same tasks (*novelty*, *divergent thinking*,



and *insight*). As such, we were unable to use tasks to represent all of the original nine constructs to create a full task-based  $9 \times 9$  similarity space of all the terms. Instead, as a proof of concept to demonstrate the viability of the approach, we built a task-based similarity matrix for six of the terms for which we could find unique task mappings: Stroop was used for *cognitive control*, Task Switching for *flexibility*, Fluency for *generation*, Mental Rotation for *imagery*, and the Wisconsin Card Sorting Task for *reasoning*.

Further, we aimed to demonstrate that the same methodology used above—in which we replace each task in succession and compare the resulting model correlations—could be applied here as well. Therefore, we used the same three tasks as above (Flanker, Go/No-Go, and Stroop) to represent *cognitive control*. And in order to test which candidate task (between Analogical Reasoning and Verb Generation) was a better fit to measure *creativity* (e.g., Beaty et al., 2017; Green et al., 2015, 2009), we used the same approach to substitute in these two tasks in successive models. This process resulted in 5 distinct neural models (i.e., similarity matrices) derived from NeuroSynth task-based meta-analyses. We then correlated these models—both full models and row-wise term correlations—between the neural data and the experts' behavioral ratings for the included terms. Results are reported in Table 4.

As with the previous analysis, results demonstrate that the choice of task greatly influences the fit between the experts' model and the neural data putatively reflecting the same constructs. An extreme example of this effect is that when Go/No-Go is used to represent *cognitive control* and Verb Generation is used for *creativity*, the full models show the worst overall fit between neural and behavioral data, correlating negatively  $r(15) = -0.52$ ,  $z = -2.00$ ,  $p = .04$ . Similarly, the two tasks used to measure *creativity* led to different results in the Stroop model: while the term-wise correlation between Analogical Reasoning (neural data) and *creativity* (behavioral data) was  $r(5) = 0.46$ ,  $z = 1.05$ ,  $p = .15$ , the correlation between Verb Generation (neural data) and *creativity* (behavioral data) was  $r(5) = -0.12$ ,  $z = -0.26$ ,  $p = .40$ . This difference highlights the critical role the task used to operationalize the cognitive constructs (e.g., *creativity*), plays in how well the neural signal reflects the putative cognitive construct in a way that is consistent with how researchers in the field conceptualize this construct.

### 3. Discussion

In order to progress towards a clearly defined understanding of the neurocognitive constituents of creative thinking, a well-defined ontology of creativity measurement is needed. Such ontological development will facilitate convergence among the scientific community on a set of constructs and operationally validated tasks that measure these constructs. The present study demonstrated a proof-of-concept for data-analytic methodology that can support the achievement of this long-term objective. Specifically, this work demonstrated how a data-driven meta-analytic approach to aggregate neuroimaging data can identify a set of experimental tasks that elicit neural activity optimally reflecting the similarity/dissimilarity of a targeted set of cognitive constructs. Evidence for the efficacy of this approach has implications for creativity research as creativity neuroscience expands the available neural data. However, there is nothing about this approach that is unique to creativity, and nearer-term value might well be gained by applying these methods to more canonical areas of psychological inquiry such as memory, executive function, and emotion, for which far more extensive neuroimaging literatures already exist.

Our analysis approach aims to quantify the degree to which a choice of experimental task will affect the fit between an observed neural response and an expected cognitive construct. As a demonstration of the methodology, we tested different neural meta-analytic models using three different tasks to measure *cognitive control* and two different tasks to measure *creativity*. In terms of the tasks that best aligned with conceptual models of their corresponding constructs, the Stroop task emerged to be the best aligned with the construct of *cognitive control* and tasks

that used an analogical reasoning paradigm were best aligned with the construct of *creativity*. While the current results should be considered exploratory (see limitations, described below), these results illustrate the type of insight that using meta-analytic representational similarity analysis can contribute toward the goal of developing an ontology of creativity.

Similarly, at the level of constructs, we found that neural meta-analyses of *flexibility* and *imagery* were the least-well aligned with the corresponding constructs within the expert-informed conceptual model. Consequently, the model fit was improved when these terms were removed. In contrast, *cognitive control*, *divergent thinking*, and *novelty* showed stronger correlations between the expert model and the neural model. Therefore, removing these terms worsens the fit between neural data and the expert conceptual model. Whereas the task-based analysis indicated which tasks elicit cognitive constructs that are well reflected in neural activity, by examining where the neural data are aligned or misaligned with the experimenters' model on the level of constructs, we can learn about how well aligned the expert conceptual model is to the neural data of the field as a whole, aggregated over numerous tasks. Such insights can drive future research in terms of examining both the neural models and the cognitive models with the overall goal of calibrating the two models to improve the fit between them.

Taken together, our results highlight how removing or adding constructs and tasks in a neural-behavioral model changes its goodness-of-fit, and how this approach can be used to study the accuracy of specific tasks for operationalizing cognitive constructs. Furthermore, building on previous research (Poldrack et al., 2011; Poldrack and Yarkoni, 2016), this work demonstrated the strength of meta-analytic neural maps in analyzing cognitive constructs. In relation to previous work with related goals and methods, some notable progress has been made in developing cognitive ontologies using behavioral data (e.g., (Poldrack et al., 2011), and a few studies have used neuroscience data to partially validate cognitive ontologies in other research areas (Eisenberg et al., 2019; Lenartowicz et al., 2010; Sabb et al., 2008, 2009). Sabb et al. (2008; 2009) applied a bibliometric analysis over PubMed to evaluate the relationship between heritability, behavior, and constructs of executive functions. Such approaches have revealed important insights regarding constructs in the executive function literature, and how these terms relate to cognitive control (Sabb et al., 2008).

Building on these efforts, Lenartowicz et al. (2010) examined whether the cognitive ontology uncovered by Sabb et al. (2008) can be mapped onto neural systems. To do so, the authors conducted a meta-analysis of brain activation across a range of tasks related to these ontological terms. This was achieved via the BrainMap database (Laird et al., 2005). These efforts and others have demonstrated the utility of comparing patterns of brain activation evoked by different cognitive tasks in order to map constructs of the mind onto structures of the brain (Lenartowicz et al., 2010; Poldrack and Yarkoni, 2016; Varoquaux et al., 2018). However, less research has explored the possibility of ontological mapping in the other direction: building a bottom-up ontology of mental constructs by starting with a data-driven, brain-based approach to explore how different tasks and sub-components of a construct relate to one another in a neurally-defined representational space. The results of our current study indicate a path that leads toward filling this gap in the literature by demonstrating that neural data and expert conceptualizations can be used together to further the ontological development of creativity measurement. Notably, this approach can also be applied more broadly to other domains of cognitive neuroscience.

Finally, it is important to emphasize again that this study is an initial proof-of-concept. Our overall goal was primarily to develop and illustrate a methodological approach that we believe has long-term potential for integrating neural data into the ontological development of creativity measurement. However, we did not seek to collect the requisite data to draw strong conclusions from the present results, and the reported analyses are constrained by several limitations. For instance, while a relative strength of NeuroSynth is the vast number of studies included

in the database, the data reported for each study are not as extensive as they could be. In particular, neural activations for each neuroimaging study are included in a meta-analysis on the level of the *publication*, rather than on the level of the *neuroimaging analysis*. This organizational structure can at times become problematic. Consider, for example, the case of a single study that includes two task conditions, one aimed at *divergent thinking* task and one aimed at *convergent thinking*. The results of these two task conditions would appear in all of the same meta-analyses, despite the fact that they presumably would show very different patterns of activation and reflect very different constructs. Such noise in the database would be attenuated by an analysis approach that operates on the level of *analyses* rather than *whole studies*, perhaps by allowing access to the original data (e.g., NeuroSynth; Poldrack and Yarkoni, 2016; Yarkoni et al., 2011) or by providing more extensive meta-data coding (e.g., BrainMap; Laird et al., 2005).

Another limitation concerns the number of available studies relevant to our focus on creativity and related tasks and constructs. Despite the fact that the NeuroSynth database contains data referring to over 14,000 studies, this number of studies still reflects only about 20–30% of the total number of neuroimaging studies conducted. Moreover, there remains only a comparatively small—though increasing—number of neuroimaging studies that have specifically focused on creativity. Thus, constructs such as *creativity*, and even related concepts such as *divergent thinking* and *convergent thinking* have relatively few studies associated with them compared to, e.g., *cognitive control*. Indeed, higher-level cognitive correlates of creativity (such as mental modeling or visuospatial reasoning) were not included in the term list as there were not sufficient data available for these constructs in the NeuroSynth database. Consequently, our terms and task list does not capture the entire space of creativity. In time, this issue will hopefully be resolved by the steady increase in the volume of creativity neuroscience studies. At present, however, due to these limitations as well as the constraints of our selection approach (described above), our analyses were limited to a smaller set of tasks and cognitive constructs than might have been ideal. Consequently, many of our neural model similarity spaces were more sparsely populated than we would have liked, and many of our correlation tests were under-powered for reaching firm conclusions about the constructs and tasks. Therefore, future research should provide a larger replication and extension of our approach, examining a much larger number of studies reflecting a more comprehensive set of terms and tasks (e.g., Eisenberg et al., 2019). Finally, even in the short term, surveying larger numbers of experts in the field and more extensive searching of experimental tasks that reflect the relevant cognitive constructs could certainly produce a more extensive space of neural and conceptual models to explore and to examine with the current methods.

These limitations notwithstanding, the present work provides a promising indication that methods such as those described here can contribute to building an ontology of measurement suitable to overcoming historical constraints and advancing understanding of human creativity. In this way, we hope that the methods described here can be useful in developing an ontology that can serve at least two major functions as the field develops: 1) converge on a set of constituent *cognitive constructs* that together—by virtue of their relations to each other—comprise a multi-dimensional representation of the complex construct of creativity; and 2) converge on a set of *experimental tasks* that reliably evoke neural activity reflective of these individual cognitive constructs. Therefore, in addition to the particular utility of these methods, they may be more broadly useful in overcoming historical constraints by helping to re-frame how researchers conceptualize and measure creativity. Instead of asking, *What is creativity?* with the expectation that a unitary construct can be satisfactorily defined, it may be more fruitful to think about creativity as a multi-dimensional similarity space and begin to optimize our tasks to measure different cognitive elements within the space of *creativity*.

## Acknowledgments

The authors wish to gratefully acknowledge funding from the National Science Foundation (US) that made this work possible; NSF-DRL #1848181 to DJMK and AEG; NSF-DRL #1661088 to DJMK; NSF-DRL #1661065 to AEG; as well as a GRFP Award to RAC.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2020.117166.

## References

- Abraham, A., 2013. The promises and perils of the neuroscience of creativity. *Front. Hum. Neurosci.* 7, 246.
- Abraham, A., 2014. Creative thinking as orchestrated by semantic processing versus cognitive control brain networks. *Front. Hum. Neurosci.* 8, 95.
- Aron, A., Aron, E.N., Smollan, D., 1992. Inclusion of other in the self scale and the structure of interpersonal closeness. *J. Pers. Soc. Psychol.* 63 (4), 596–612.
- Barbot, B., Hass, R.W., Reiter-Palmon, R., 2019. Creativity assessment in psychological research: (Re) setting the standards. *Psychol. Aesthet. Creat. Arts* 13 (2), 233.
- Beatty, R.E., Christensen, A.P., Benedek, M., Silvia, P.J., Schacter, D.L., 2017. Creative constraints: Brain activity and network dynamics underlying semantic interference during idea production. *Neuroimage* 148, 189–196.
- Beatty, R.E., Seli, P., Schacter, D.L., 2019. Network neuroscience of creative cognition: mapping cognitive mechanisms and individual differences in the creative brain. *Curr. Opin. Behav. Sci.* 27, 22–30.
- Benedek, M., Fink, A., 2019. Toward a neurocognitive framework of creative cognition: the role of memory, attention, and cognitive control. *Curr. Opin. Behav. Sci.* 27, 1116–1122.
- Bilder, R.M., Sabb, F.W., Parker, D.S., Kalar, D., Chu, W.W., Fox, J., et al., 2009. Cognitive ontologies for neuropsychiatric phenomics research. *Cognit. Neuropsychiatry* 14 (4–5), 419–450.
- Chen, C., Yang, J., Lai, J., Li, H., Yuan, J., 2015. Correlating gray matter volume with individual difference in the flanker interference effect. *PLoS One* 10 (8), e0136877.
- Cortes, R.A., Weinberger, A.B., Daker, R.J., Green, A.E., 2019. Re-examining prominent measures of divergent and convergent creativity. *Curr. Opin. Behav. Sci.* 27, 90–93.
- Council, N.R., Committee, G.S., 2005. Learning to Think Spatially. National Academies Press.
- Dietrich, A., 2007. Who's afraid of a cognitive neuroscience of creativity. *Methods* 42 (1), 22–27.
- Dietrich, A., 2019. Where in the brain is creativity: a brief account of a wild-goose chase. *Curr. Opin. Behav. Sci.* 27, 36–39.
- Eisenberg, I.W., Bissett, P.G., Enkavi, A.Z., Li, J., MacKinnon, D.P., Marsch, L.A., et al., 2019. Uncovering the structure of self-regulation through data-driven ontology discovery. *Nat. Commun.* 10 (1), 2319.
- Eklund, A., Nichols, T.E., Knutsson, H., 2016. Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proc. Natl. Acad. Sci.* 113 (28), 7900–7905.
- Florida, R., 2014. The creative class and economic development. *Econ. Dev. Q.* 28 (3), 196–205.
- Frey, C.B., Osborne, M.A., 2017. The future of employment: How susceptible are jobs to computerisation. *Technol. Forecast. Soc. Change* 114, 254–280.
- Glăveanu, V.P., Hanchett Hanson, M., Baer, J., Barbot, B., Clapp, E.P., Corazza, G.E., et al., 2019. Advancing creativity theory and research: A socio-cultural manifesto. *J. Creat. Behav.*
- Glăveanu, V.P., Kaufman, J.C., 2019. The creativity matrix: Spotlights and blind spots in our understanding of the phenomenon. *J. Creat. Behav.*
- Green, A.E., Cohen, M.S., Raab, H.A., Yedibalian, C.G., Gray, J.R., 2015. Frontopolar activity and connectivity support dynamic conscious augmentation of creative state. *Hum. Brain Mapp.* 36 (3), 923–934.
- Green, A.E., Kraemer, D.J.M., Fugelsang, J.A., Gray, J.R., Dunbar, K.N., 2009. Connecting long distance: semantic distance in analogical reasoning modulates frontopolar cortex activity. *Cereb. Cortex* 20 (1), 70–76.
- Guilford, J.P., 1950. Creativity. *Am. Psychol.* 5.
- Jennings, K.E., 2010. Developing creativity: Artificial barriers in artificial intelligence. *Minds Mach.* 20 (4), 489–501.
- Kriegeskorte, N., Mur, M., Bandettini, P.A., 2008. Representational similarity analysis - connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* 2, 4.
- Laird, A.R., Lancaster, J.J., Fox, P.T., 2005. Brainmap. *Neuroinformatics* 3 (1), 65–77.
- Lenartowicz, A., Kalar, D.J., Congdon, E., Poldrack, R.A., 2010. Towards an ontology of cognitive control. *Top. Cognit. Sci.* 2 (4), 678–692.
- Lichtenberg, J., Woock, C., & Wright, M. (2008). *Ready to innovate: Are educators and executives aligned on the creative readiness of the US workforce?*
- Liu, C., Chen, Z., Wang, T., Tang, D., Hitchman, G., Sun, J., et al., 2015. Predicting stroop effect from spontaneous neuronal activity: a study of regional homogeneity. *PLoS One* 10 (5), e0124405.

- McCormick, E.M., Qu, Y., Telzer, E.H., 2016. Adolescent neurodevelopment of cognitive control and risk-taking in negative family contexts. *Neuroimage* 124, 989–996.
- Necka, E.A., Sokolowski, H.M., Lyons, I.M., 2015. The role of self-math overlap in understanding math anxiety and the relation between math anxiety and performance. *Front. Psychol.* 6, 1543.
- Newcombe, N. (2017). *Harnessing spatial thinking to support stem learning*.
- Piffer, D., 2012. Can creativity be measured? An attempt to clarify the notion of creativity and general directions for future research. *Think. Skills Creat.* 7 (3), 258–264.
- Poldrack, R.A., Kittur, A., Kalar, D., Miller, E., Seppa, C., Gil, Y., et al., 2011. The cognitive atlas: toward a knowledge foundation for cognitive neuroscience. *Front. Neuroinform.* 5, 17.
- Poldrack, R.A., Yarkoni, T., 2016. From brain maps to cognitive ontologies: informatics and the search for mental structure. *Annu. Rev. Psychol.* 67, 587–612.
- Runco, M.A., Jaeger, G.J., 2012. The standard definition of creativity. *Creat. Res. J.* 24 (1), 92–96.
- Sabb, F.W., Bearden, C.E., Glahn, D.C., Parker, D.S., Freimer, N.B., Bilder, R.M., 2008. A collaborative knowledge base for cognitive phenomics. *Mol. Psychiatry* 13 (4), 350–360.
- Sabb, F.W., Burggren, A.C., Higier, R.G., Fox, J., He, J., Parker, D.S., et al., 2009. Challenges in phenotype definition in the whole-genome era: multivariate models of memory and intelligence. *Neuroscience* 164 (1), 88–107.
- Shin, G., Kim, C., 2015. Neural correlates of cognitive style and flexible cognitive control. *Neuroimage* 113, 78–85.
- Varoquaux, G., Schwartz, Y., Poldrack, R.A., Gauthier, B., Bzdok, D., Poline, J.-B., et al., 2018. Atlases of cognition with large-scale human brain mapping. *PLoS Comput. Biol.* 14 (11), e1006565.
- Grajewska, vel, Ž., B., Sim, E.-J., Hoenig, K., Herrnberger, B., Kiefer, M., 2011. Mechanisms underlying flexible adaptation of cognitive control: Behavioral and neuroimaging evidence in a flanker task. *Brain Res.* 1421, 52–65.
- Wager, T.D., Sylvester, C.-Y.C., Lacey, S.C., Nee, D.E., Franklin, M., Jonides, J., 2005. Common and unique components of response inhibition revealed by fMRI. *Neuroimage* 27 (2), 323–340.
- Ward, T.B., 2007. Creative cognition as a window on creativity. *Methods* 42, 28–37.
- Ward, T.B., Smith, S.M., Finke, R.A., 1999. Creative cognition. In: Sternberg, R.J. (Ed.), *Handbook of Creativity*. Cambridge University Press, New York, pp. 189–212.
- Yarkoni, T., Poldrack, R.A., Nichols, T.E., Van Essen, D.C., Wager, T.D., 2011. Large-scale automated synthesis of human functional neuroimaging data. *Nat. Methods* 8 (8), 665–670.