COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 42 (2018) 885–917

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12569

# Coherence in the Visual Imagination

Michael O. Vertolli, Matthew A. Kelly, Jim Davies

*Institute of Cognitive Science, Carleton University*

Received 9 February 2015; received in revised form 8 May 2017; accepted 2 October 2017

**Abstract**

An incoherent visualization is when aspects of different senses of a word (e.g., the biological "mouse" vs. the computer "mouse") are present in the same visualization (e.g., a visualization of a biological mouse in the same image with a computer tower). We describe and implement a new model of creating contextual coherence in the visual imagination called Coherencer, based on the SOILIE model of imagination. We show that Coherencer is able to generate scene descriptions that are more coherent than SOILIE's original approach as well as a parallel connectionist algorithm that is considered competitive in the literature on general coherence. We also show that co-occurrence probabilities are a better association representation than holographic vectors and that better models of coherence improve the resulting output independent of the association type that is used. Theoretically, we show that Coherencer is consistent with other models of cognitive generation. In particular, Coherencer is a similar, but more cognitively plausible model than the $C^3$ model of concept combination created by Costello and Keane (2000). We show that Coherencer is also consistent with both the modal schematic indices of perceptual symbol systems theory (Barsalou, 1999) and the amodal contextual constraints of Thagard's (2002) theory of coherence. Finally, we describe how Coherencer is consistent with contemporary research on the hippocampus, and we show evidence that the process of making a visualization coherent is serial.

*Keywords:* Imagination; Coherence; Visualization; Cognitive modeling; Hippocampus

## 1. Introduction

The imagination is implicated in many aspects of human cognition, including planning, problem solving, hypothetical thinking, counterfactual thinking, theory of mind, and mental time travel (Davies, Atance, & Martin-Ordas, 2011). Despite extensive research on imagination as a facilitator for these abilities (see, e.g., Markman, Klein, & Suhr, 2012), and on the properties of mental images (Kosslyn, 1996), the processes that generate the content of imagined scenes are largely unstudied.

Correspondence should be sent to Michael O. Vertolli, Institute of Cognitive Science, Carleton University, 1125 Colonel By Drive, Ottawa, ON, K1S 5B6 Canada. E-mail: michaelvertolli@gmail.com

This work focuses on imagination related to the most studied sensory modality: vision. When someone imagines a visual scene (e.g., of a mouse eating cheese), she can use visual memories from many different experiences as the components of the new scene. How these components are selected from memory is not obvious. If more than the mouse and cheese are included in the scene, it is unclear what makes the selection of some elements (e.g., a cat, mousetrap, floorboards, or a countertop) more likely than others (e.g., a rollercoaster, map of Belarus, or cruise ship).

What is known is that people do not arbitrarily select the components for their imaginings, even if those imaginings are entirely fictional (Cockbain, Vertolli, & Davies, 2014; Ward, 1994). There is an intuitive coherence imposed on imagined scenes that inhibits unusual and sometimes even highly creative combinations. Our mental images, for the most part, make sense and are not populated by random objects in random placements. How objects are selected such that imagined scenes are coherent is an important subproblem of the generative imagination task.[1]

Visual perception takes information from the world and, through different processes, detects increasingly abstract properties (Attneave, 1954; Barlow, 1961). Unlike vision, imagination must use stored representations to construct an imagined scene (e.g., a mouse eating cheese). Imagination goes beyond what is explicitly represented in the text or memory, generating novel combinations of imagined things, in novel arrangements.[2] At the same time, these novel combinations must be coherent and realistic—that is, they must conform to the regularities of the world as it has been experienced. A mouse and cheese belong together in a way that a mouse and a space shuttle do not. How the mind creates coherent, novel scenes in imagination is the subject of this research.

Mental imagery—the rendering of an internal, depictive image that resembles what one would see in the real world (Kosslyn, 1996)—is often thought to be identical with visual imagination. However, we see the rendering of a mental image as a final, optional stage in the overall generative process. Coherence relations, in this view, are determined in cognitive processing prior to the low-level depictive parameters of mental imagery. That is, if visual mental imagery exists (and we acknowledge that it is a debate in the literature; see Kosslyn, Thompson, & Ganis, 2006; Pylyshyn, 2002), it requires some kind of previously generated scene description to render as colors at particular locations. It is worth qualifying that we are currently agnostic to the particulars of the instantiation of this scene description. In sum, our goal is to create a model of coherence generation that is psychologically and neurally plausible. We present an implemented cognitive model, called Coherencer, and compare it to three other models: an approximation of a naive Bayesian network, a connectionist model, and a holographic model.

In each case, we expect that Coherencer will outperform the other models. Each of the functional distinctions between it and the other models provides specific and important advantages to the coherence process. We discuss these distinctions before each of the comparisons.

## 2. Models

In this section, we address each of the models in turn. We begin with a detailed discussion of the original model that preceded Coherencer and the larger architecture of which it is a part. We then consider our own implementation of Thagard's connectionist algorithm. Finally, we address the holographic memory model.

### *2.1. Coherencer*

The Science of Imagination Laboratory Imagination Engine (SOILIE) is a computational model of the generative processes of imagination (Breault, Ouellet, Somers, & Davies, 2013). SOILIE takes a single word as input (the "query") and returns a collection of associated labels with their relative positions on a two-dimensional mental canvas. SOILIE's processes are designed to model the way humans create visual imagined scene descriptions. The resulting scenes can be rendered to more easily assess the output, but this rendering process is not designed to model human behavior.

The core theory behind SOILIE is that novel imagined scenes are recombinations of elements from previous experience. SOILIE's "experiences" are labeled images from the web. SOILIE must determine which labels are appropriate to select, given a particular query, when generating a scene. SOILIE uses co-occurrence relations to make this selection. The motivation for this is that if an agent is tasked with imagining a scene with, say, a "car" in it, then it will choose other objects in the scene based on what objects have appeared with cars in memory. In the implementation, co-occurrence is determined by the frequency with which one label is present in the same image with another label. We acknowledge that human beings use more complex structures to determine what goes in visual scenes (e.g., schemas), but our goal with this work is to understand how simple associations are used.

SOILIE uses the Peekaboom database of labeled images as a substitute for human visual memory. The dataset is the combined result of two online games: the ESP Game and Peekaboom (Von Ahn & Dabbish, 2004; Von Ahn, Liu, & Blum, 2006). This database consists of approximately 50,000 images, with an average of 12 labels each. These labels are associated with pixel locations in the images of the objects these labels describe. Thus, with this database, SOILIE knows *what* is in each image, and *where* in the image those objects are.

SOILIE derives co-occurrence probabilities from the conditional relative frequencies of labels in the Peekaboom database. Co-occurrence probabilities are calculated by dividing the total number of images ($I$) in the Peekaboom database that contain the co-occurring label ($l$) *and* a particular query ($q$) by the total number of images with just the query. Using set theory notation, this yields:

$$P(l|q) = \frac{|I_q \cap I_l|}{|I_q|} \tag{1}$$

where ∩ indicates set intersection and ‖ indicates cardinality (i.e., the total number of elements in the set). One important feature of this formalization is that it is non-commutative (i.e., $P$ yields a different probability for mouse given cheese than it does for cheese given mouse). Parallel research on co-occurrence in machine learning suggests that although this is more realistic (e.g., almost all weddings have flowers but most flowers are not in weddings), most models do not account for it (Huang, Yu, & Zhou, 2012; Zhang & Zhou, 2013).

Research in neuroscience suggests that visual working memory can hold approximately three to five chunks of "average complexity" (Cowan, 2001; Edin et al., 2009). As such, SOILIE retrieves four new labels, in addition to the query, to be placed in the imagined scene. It is possible that chunking (i.e., combining two or more elements) can occur, but we chose to ignore it for simplicity. The result is that four labels, excluding the query, are retrieved by SOILIE from the co-occurrence data.

### 2.1.1. *The problem: Incoherence*

After working with earlier instantiations of SOILIE (Breault et al., 2013), a problem became apparent. When images are selected using labels with the highest co-occurrence, or the "top-*n*," labels,[3] the scenes produced are often contextually incoherent. For example, SOILIE was queried with the word "mouse," which is polysemous (i.e., it has multiple, related meanings; in this case, a computer mouse and the animal mouse). SOILIE returned an image containing "animal," "computer," and "monitor," elements from different senses of the word in the same image as a result of the underlying polysemy. We call this problem "incoherence."

Reducing images to co-occurrence probabilities in visual memory, as SOILIE does, makes the images, labels, and co-occurrence relations that separate the polysemous meanings no longer directly detectable. They are collapsed into a single dimension associating pairs of labels (see Table 1).

Incoherence and related problems caused by dimensionality collapse are not limited to word overlap and polysemy. Given a particular selection of labels, all of which co-occur, it is still possible that no single image exists in the database that has all of those labels. For example, one image might contain "doctor," "needle," "policeman," and "gun"; another

Table 1
Label co-occurrence probability of two images alone and in SOILIE's complete database. This shows how incoherent scene descriptions can emerge when looking at the top co-occurring labels

**Image 1 labels:** mouse, eye, rodent, rat, animal, ear, ears
**Image 2 labels:** mouse, wires, monitor, screen, headphones, computer
**Co-occurrence of each label with query "mouse" given only those two images:** 0.5
**Co-occurrence of label with query "mouse" using all images in the database:**

| | | | | | |
|---|---|---|---|---|---|
| Rat | 0.29 | Monitor | 0.12 | Eye | 0.06 |
| Ear | 0.19 | Screen | 0.10 | Headphones | 0.01 |
| Computer | 0.17 | Rodent | 0.08 | Wires | 0.01 |
| Animal | 0.13 | Ears | 0.07 | | |

**Top-4 labels for the query "mouse" based on all images using co-occurrence:**
rat, ear, computer, animal

Fig. 1. Incoherent image that blends two contexts from query "mouse" with output "rat," "ear," "computer," and "keyboard." Note that the image actually shows a groundhog that was mislabeled as a "rat."

might contain "doctor," "morphine," "surfer," and "trunks"; and a third might contain "needle," "morphine," "policeman," "gun," "surfer," and "trunks." Thus, all seven labels would co-occur in the database in general without any individual image containing the entire set. To combine the entire set would require an insensitivity to the underlying dependency relations (e.g., needle and drug to policeman in the absence or presence of a doctor). In a given context, some objects are dependent and some are not. Assuming that they are all independent is problematic, as we see in the resulting incoherent combinations.

Models that use only a single dimension, co-occurrence with the query (like SOILIE's top-$n$ model), assume the labels are conditionally independent of one another, and as such, they are unable to infer the appropriate contextual relations from differences in the underlying images. They often produce incoherent images relative to what they know about the world (i.e., the database of images; see Fig. 1).[4]

### 2.1.2. The solution: Coherencer

To solve the incoherence problem, we chose to augment the top-$n$ approach with a paired association search using a serial, local-hill searching algorithm. The resulting model is Coherencer.

Coherencer operates as follows (see Fig. 2; for a formal description, see Vertolli & Davies, 2014). First, a top-$n$ search gathers the top four co-occurring labels with the query (any co-occurrence greater than zero). Then, an associative search checks whether each label in the pool co-occurs with all the others, as well as with the query. The mean of these co-occurrence relations is the overall co-occurrence score for that list of labels. Labels with low co-occurrence in the network at a given time step are removed from the set (i.e., rejected). New labels that co-occur with the query

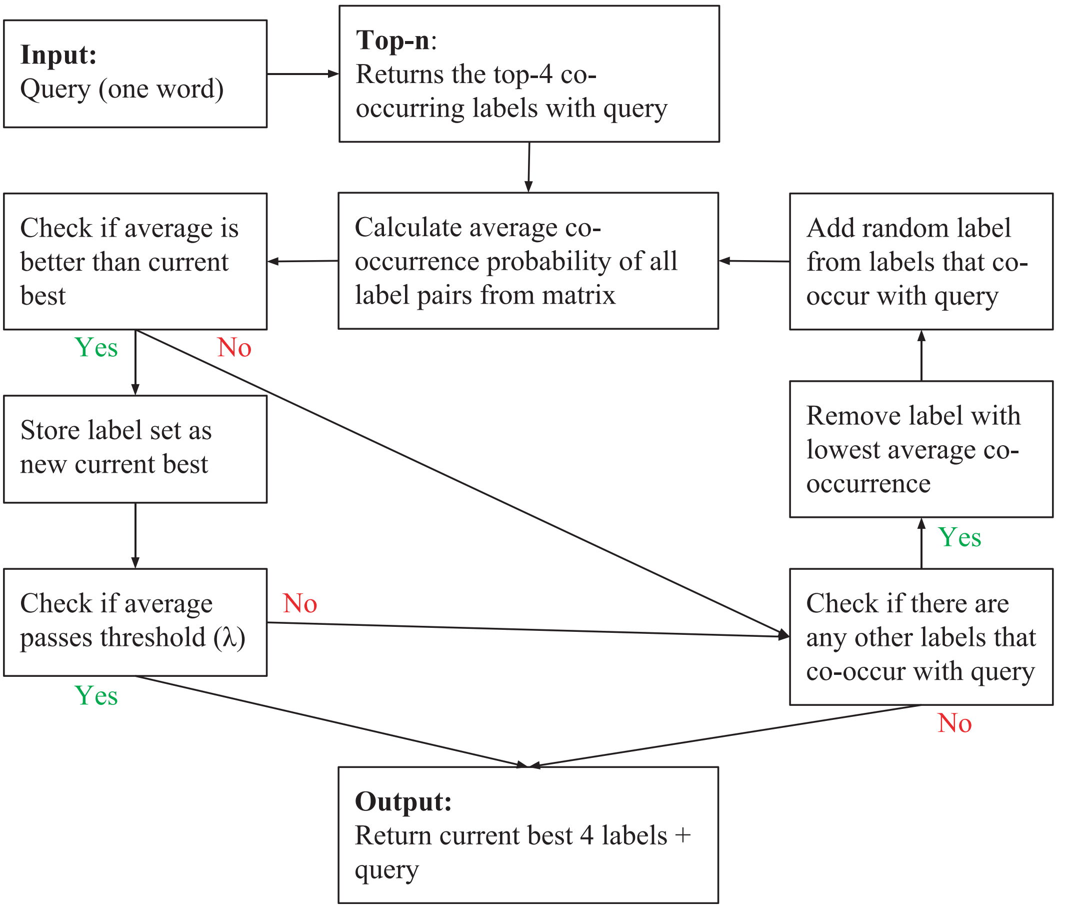


Fig. 2. Control in Coherencer.

are then randomly swapped in to replace the rejected label. This cycle repeats until the overall co-occurrence score for the list exceeds a given threshold, which is selected via a random parameter search (i.e., the threshold which achieves the best score on the validation set of images). Once the threshold is exceeded, the set that remains is returned for inclusion in the imagined scene (to be placed in locations and perhaps to be rendered later with mental imagery). If the set of labels that co-occur with the query is exhausted, the model returns the collection of labels with the highest mean co-occurrence probability.

### 2.2. *Thagard's connectionist model*

We built a connectionist model as described in Thagard (2002) in order to compare Coherencer to a locally parallel algorithm that was competitive in the coherence literature (see Fig. 3). The model was built as follows.

A node is created for the query and every label co-occurring with the query. Co-occurring labels create positive constraints where the presence of a label in an image increases the likelihood of another label in the image. For every positive constraint between two labels, an excitatory link is placed between the corresponding nodes with a weight equal to the co-occurrence probability.[5] When two words never co-occur in the memory (a negative constraint), an inhibitory connection is set between corresponding nodes with a weight set to the average of all non-zero co-occurrence probabilities, $P_{mean}$.[6] An initial activation of 0.24 is assigned to each node except for the query node, which is locked to an activation of 1.0. All of these parameters were assigned using a random parameter search. Node activations are updated until the amount of change is lower than a set threshold. All nodes (except for the query node) have their activation updated in parallel using the following formula:[7]

$$\mathbf{a}_{t+1} = \mathbf{a}_t(1 - d) + f(\mathbf{net}) \tag{2}$$

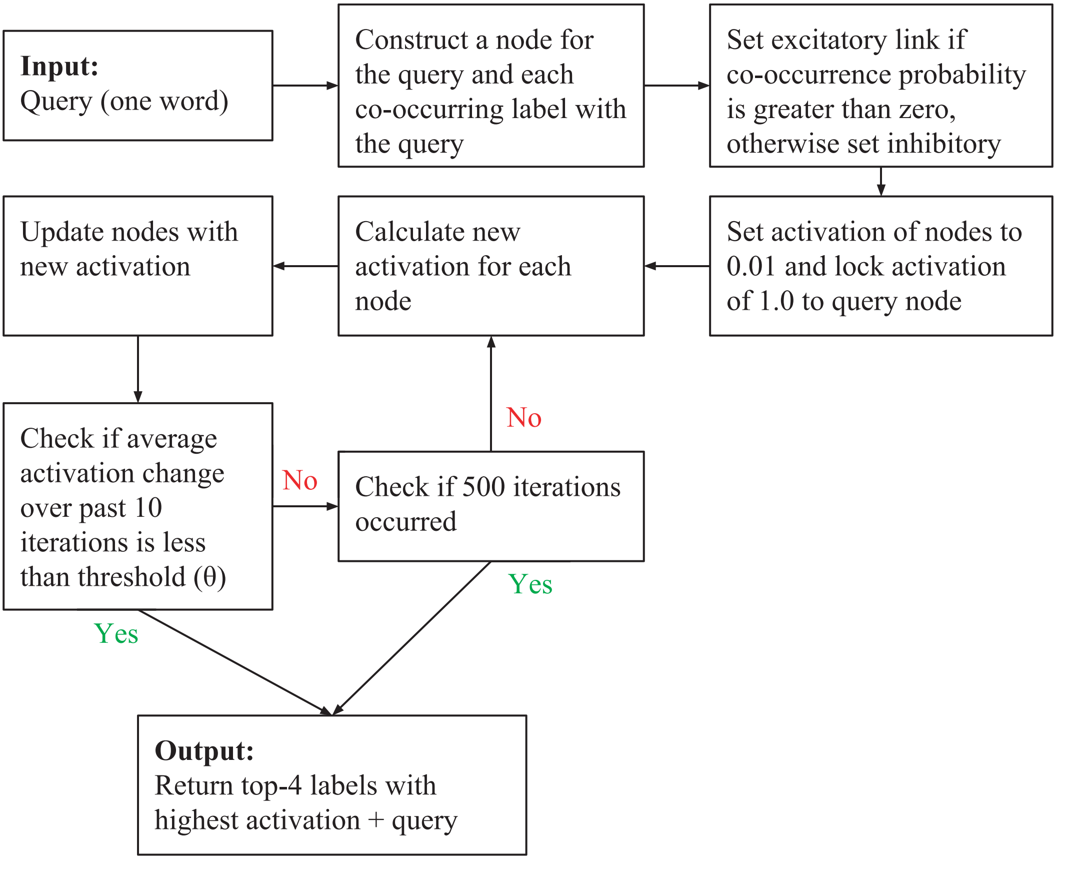


Fig. 3. Control in Thagard's model.

where $\mathbf{a}_t$ is a vector of all the node activations at time $t$, $d$ is a scalar decay parameter (0.35) that decrements each node at every cycle, and the vector **net** is computed by:

$$\mathbf{net} = \mathbf{a}_t W \tag{3}$$

where $W$ is the weight matrix for the network with its rows corresponding to the node being updated and the columns corresponding to the linked nodes (i.e., neighbors of node $i$). The values at $W_{i,i}$ (i.e., the diagonal of the matrix) are set to 0 so the activation passed from a node to itself is 0. $W$ also corresponds to Coherencer's co-occurrence matrix with all co-occurrence values of 0 set to $P_{mean}$. Finally, $f$ is a function that performs element-wise multiplication with a different number depending on the sign of each element of **net**:

$$f(\mathbf{net}) = net_i x \begin{cases} x = a_{\max} - a_i & \text{if } net_i > 0 \\ x = a_i - a_{\min} & \text{if } net_i \leq 0 \end{cases} \tag{4}$$

where $x$ is the variable multiplier, $a_i$ is the ith value of $\mathbf{a}$, $a_{\max}$ is the maximum activation of a node (1.0), and $a_{\min}$ is the minimum activation ($-1.0$). After the update, each node is constrained to the maximum and minimum activation values if it exceeds them.

In the larger process, the activations will update until the average change in the sum of all differences is less than a threshold ($\theta$) or until 261 iterations occur. The following equation illustrates the former:

$$\Delta \mathbf{a}_t = \frac{1}{10n} \sum_{t-10}^{t} \sum_{i=0}^{n} (|a_{t,i} - a_{t-1,i}|) < \theta \tag{5}$$

where $\Delta a$ is the change in activation over the past 10 iterations, $a_{t,i}$ means activation at time $t$ and node $i$, || here indicates absolute value, and the threshold $\theta$ is 0.007.[8] The four labels with the highest activation are selected for inclusion in the imagined scene, implementing a kind of top-4 filter.

## 2.3. Holographic vectors

Previously, we showed the differences between Coherencer and the top-$n$ model in terms of co-occurrence probabilities. Now, we will compare these same two models using holographic vector representations, which use a different kind of association and comparison metrics. Holographic vectors have been used to specifically handle context-based information in text (Jones & Mewhort, 2007). Although tested in a slightly different domain, it is likely to capture more salient context information than co-occurrence probabilities.

In addition to this, vector representations are considered to be a neurally plausible abstraction. They are currently used in the Semantic Pointer Architecture (SPA;

Eliasmith, 2013), which is the underlying architecture of SPAUN, currently the world's largest functional, spiking neural network model of the brain (Eliasmith et al., 2012). They have also been used in related neural simulations of creativity (Thagard & Stewart, 2011). Thus, holographic vector representations are not only mathematical abstractions of associations but can also be used for neuron-level implementations of cognitive models.

The holographic representation is set up as follows. Each label in the Peekaboom image database is represented by a vector of 1,000 dimensions.[9] Each of these vectors is generated randomly by sampling 1,000 values from a normal distribution with a mean of zero and a standard deviation of $1/\sqrt{1,000}$. As a result, each vector has a Euclidean length of approximately one and is approximately orthogonal to every other vector. These vectors are the *environment vectors* (Jones & Mewhort, 2007).

Each label is also represented by a second type of vector, termed a memory vector. A memory vector is a representation of the associations between a given label and all other labels. Memory vectors can be constructed in a number of ways. Here, we adopt the simplest of the methods that were used by Jones and Mewhort (2007). For our purposes, an image is a collection of co-occurring labels. The memory vector for a given label is the sum (using vector addition) of all environment vectors representing labels that the given label co-occurs with. If a label (A) co-occurs with the given label (B) in only one image, that label (A)'s environment vector is added to the memory vector of the given label (B) only once. If a label (A) co-occurs with the given label (B) in more than one image, that label (A) is added to the memory vector (B) once for each of those images.

In vector space models, similarity of concepts is measured as the angle, typically the cosine, between vectors that represent those concepts. The cosine ranges from 1 to $-1$. When comparing a pair of vectors, a cosine of 1 indicates that the pair of vectors is at a $0^\circ$ angle and they are identical representations. A cosine of 0 indicates that the pair is at a $90^\circ$ angle, and they are completely unrelated representations.

Effectively, the cosine between a memory vector for a label $q$ and the environment vector for another label $l$ is a noisy estimate of the label's co-occurrence probability, $P(l|q)$, for any label $l \neq q$. A given label's memory vector will be most similar to the environment vectors that represent labels that co-occurred with the given label most frequently. However, this is a noisy estimate of $P(l|q)$, and as such will be less accurate than storing the exact co-occurrence probabilities. There is no particular advantage to using vectors in this way when one can instead use exact co-occurrence probabilities.

However, holographic vectors are beneficial because of (1) their ability to represent arbitrarily complex associations in a compressed form and (2) their spatial nature, allowing for easy comparisons between any two representations to be made by measuring the cosine of the angles between them. They are also more neurally plausible (3). One can take advantage of (2), the spatial nature of the vectors. The cosine of a memory vector with another memory vector indicates how often the labels appear in *similar* images in addition to the same images. Consequently, we chose to use this representation for our comparison.

This completes our discussion of each of the models that will be compared. Each model uses a different method to take as input a single label and return four other labels that co-occur with that label in a given database. In the next section, we discuss the specific details of the experimental analysis.

## 3. Experimental comparisons

In this section, we outline the details of the experimental comparisons and their results. We begin with a discussion of the preprocessing we performed on the Peekaboom database of images and their corresponding labels as well as how we trained the models, selected their parameters, and assessed them. We also discuss a statistical metric that allows us to ensure that the differences found between the models, many of which are stochastic, are meaningful differences. We compare Coherencer to the top-*n* model and Thagard's connectionist model in the first two experiments, respectively. In the third experiment, we compare Coherencer and the top-*n* across the two association metrics: co-occurrence probabilities and holographic vectors.

Each model, including the original top-*n* model that Coherencer augments, entails a different set of general assumptions for the imagination mechanism when it is compared to Coherencer. Our evaluation of the original top-*n* model (relative to Coherencer) addresses whether imagination can assume statistical independence across its associations (i.e., the probability of one label occurring is not related to the probability of another label occurring). Our evaluation of the connectionist model deals with whether a locally serial process or a locally parallel process better captures coherence. And finally, our evaluation of the holographic model addresses how co-occurrence is represented by the mechanism: average co-occurrence probabilities or cosines between vectors.

We expect that humans can successfully infer scene-level associations with excellent fidelity. Since none of the models at present can achieve expected human scores, we argue that it is too early in this research program to be overly concerned with modeling human data. That is, we do not need human data to know that none of these models do this task as well as people do. We can productively evaluate them by comparing them to each other on a measurement of accuracy, and we do not yet need to compare them to quantitative human data—which in any case does not yet exist. When models are efficacious enough that they begin to approach human-level skill, it will then be productive to compare them to empirical human data.

The basic task is as follows: A single label is input as the query to the model. The model must then select four other labels that, together with the query, describe objects that would appear in a plausible scene.

### 3.1. Preprocessing

The entire Peekaboom database was filtered to remove all images with fewer than five labels, and any labels that only occurred in those images (e.g., if "platypus" occurred in images with fewer than five labels, and in no images with more than five labels, "platypus" would be deleted from the database).

The database was then divided into a training, validation, and test sets in proportions of 50-16-33, respectively. Images and labels were then recursively removed until every image had at least five labels and every label was in each of the three sets. A total of 2,032 labels and 18,947 images remained after filtration. The training set of images was then compressed to a data structure containing only the co-occurrence of labels expressed as one of the association metrics: co-occurrence probabilities (Eq. 1) or the cosine scores between holographic vectors. This was then input to each of the systems.

The parameters of each model were selected via a random parameter search on the validation set of images, which has been shown to be a competitive approach in the literature (Bergstra & Bengio, 2012). The results of the parameter search for Coherencer and Thagard's model were very stable, so the search was terminated after 38 and 111 iterations, respectively. For the holographic vector representation, the results were consistently higher with the higher cosine thresholds, so we selected the highest possible threshold (1.0) in an effort to maximize this trend. In all cases, scores were consistent across validation and test sets.

Since a cosine threshold of 1.0 is unusually high, we checked whether or not Coherencer was actually able to find label combinations that met that threshold in the holographic case. On every label simulated, Coherencer did not find *any* label sets at that threshold. Instead, the model would search through the entire set of co-occurring labels with the query and then return the best set it found over the course of that search. Although this was an interesting finding, its exploration was left to future work.

### 3.2. Model run metric

The number of model runs used in each experiment conforms to Byrne's (2013) analytic model run metric. The metric specifies the necessary number of runs for a robust result. It is derived from the formula for confidence intervals for proportions using the following formula:

$$n = p(1-p)\left(\frac{z}{w}\right)^2 \tag{6}$$

where $n$ is the necessary number of model runs, $p$ is the proportion of successes for the model, $z$ is the standard score of the desired confidence interval distribution, and $w$ is half the difference between the two score proportions (e.g., comparing 20% success rate to 30%, half the difference is 10/2 = 5% and the $w$ value is 0.05). The proportional success rate for this calculation is based on a single trial run of each model. Effectively, the model run metric reverses the standard confidence interval equation to figure out the number of runs necessary ($n$) to achieve that level of confidence ($z$) given the size of difference between the results of the two models compared ($w$) and the results of each of the models alone ($p$).

### 3.3. Coherencer and top-n

The first comparison is between Coherencer and the original, top-$n$ model. If at least one of the original images contains all five labels that a model outputs for the generated image, including the query, the model scores a point. If no images contain all five labels, no point is scored. The total points scored by a model are compared, *where the other model failed to score a point* for the same query (i.e., excluding labels where both models failed or both models succeeded).

Each model was run using all 2,032 labels as queries. The top-$n$ procedure always yields the same result (the top-4 associated labels), so it was run once per query label. Coherencer has stochastic variation in its results, so it was run 83 times on all 2,032 labels based on Byrne's (2013) metric. The totals were averaged across runs. We expected Coherencer to outperform the top-$n$ model.

#### 3.3.1. Results

Coherencer had more successful matches than the top-$n$ model, as hypothesized:[10] McNemar's repeated measures chi-square test demonstrates that Coherencer performed significantly better than top-$n$, $\chi^2(1, N = 2{,}032) = 239.00$, $p < .000$, $\phi = .20$. The average scores in each of the categories are listed in Table 2. Model runs where Coherencer and top-$n$ both fail or both succeed on a given query (i.e., the models perform identically) are ignored. The comparison occurs on runs where one model failed and the other succeeded. All values are reported for completeness. Both the actual number of runs and the statistically expected number of runs for a given category are reported (see Table 2).

### 3.4. Coherencer and Thagard's model

In this section, we discuss the comparison between Coherencer and Thagard's model. The method is the same as that used to compare Coherencer and top-$n$. To reiterate, a

Table 2
McNemar $\chi^2$ calculation between Coherencer and top-$n$

| | | Coherencer Failure | Coherencer Success | Total |
|---|---|---|---|---|
| Top-$n$ failure | Actual | 1,272.0 | 498.0 | 1,770.0 |
| | Expected | 1,208.2 | 561.8 | |
| Top-$n$ success | Actual | 115.0 | 147.0 | 262.0 |
| | Expected | 178.8 | 83.2 | |
| Total | | 1,387.0 | 645.0 | 2,032.0 |

Table 3
McNemar $\chi^2$ calculation between Coherencer and Thagard's model (TM)

| | | Coherencer Failure | Coherencer Success | Total |
|---|---|---|---|---|
| TM failure | Actual | 1,370.0 | 620.0 | 1,990.0 |
| | Expected | 1,358.3 | 631.7 | |
| TM success | Actual | 17.0 | 25.0 | 42.0 |
| | Expected | 28.7 | 13.3 | |
| Total | | 1,387.0 | 645.0 | 2,032.0 |

model successfully describes an imagined scene when it is realistic, as determined by whether there was an image in the test set that had the same set of labels as the imagined scene. Byrne's (2013) metric dictates 41 runs for each model. We expected Coherencer to outperform Thagard's model.

#### *3.4.1. Results*

Coherencer had more successful matches than Thagard's model, as hypothesized: McNemar's repeated measures chi-square test demonstrates that Coherencer performed significantly better than Thagard's model, $\chi^2(1,\ N = 2{,}032) = 548.00$, $p < .000$, $\phi = .09$. The average scores in each of the categories are listed in Table 3. Model runs where Coherencer and Thagard's model both fail or both succeed on a given query are ignored. The comparison occurs between the runs where one model failed and the other succeeded. As is standard with chi-square tests, both the actual number of runs and the statistically expected number of runs for a given category are reported.

### *3.5. Co-occurrence and holographic*

We discuss the comparison between co-occurrence probabilities and holographic vector representations of association in terms of the top-*n* and Coherencer models. This means that four model-representation pairs were compared: co-occurrence-top-*n*, co-occurrence-Coherencer, holographic-top-*n*, and holographic-Coherencer. The current method follows the same outline as the previous two. However, due to the complexity of the comparison across four conditions, the results could no longer be paired for comparison. As a consequence, the total successes and failures of each condition were compared as a whole. Byrne's (2013) metric required 83 model runs for Coherencer on the conditions using co-occurrence probabilities and 130 runs on the holographic vector conditions. Top-*n* was run once for each type of association. The hypothesis was that Coherencer would outperform the top-*n* model across both association types, following the results of the initial experimental comparison. We also expected the holographic memory vector associations, by capturing more contextual data, would outperform the co-occurrence probability representation across both models.

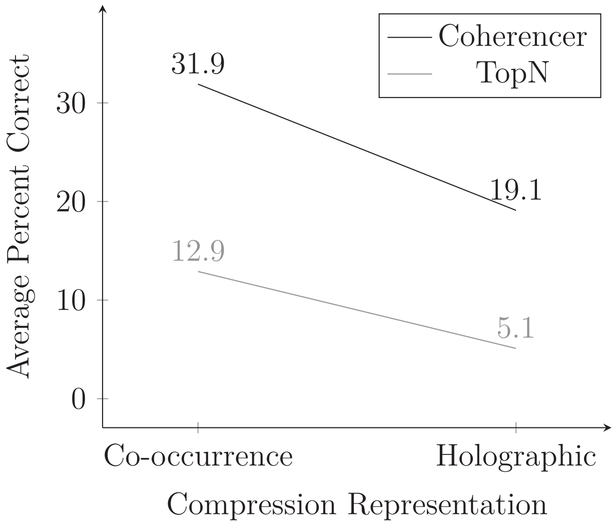


Fig. 4. Model success rate with different association type-model pairs out of a total 2,032 possible successes. All error bars were omitted as they were less than 0.01% across all of Coherencer's conditions.

*3.5.1. Results*

Coherencer outperformed the top-*n* model across both types of association, as hypothesized, replicating our initial comparison. Contrary to our hypothesis, the co-occurrence probability representation outperformed the holographic vector representation across both models. The success rates out of the 2,032 possible query labels for each of the four conditions (top-*n* and co-occurrence, top-*n* and holographic, Coherencer and co-occurrence, Coherencer and holographic) are shown in Fig. 4.

As a consequence of the categorical nature of the two independent variables (e.g., top-*n* or Coherencer) and the dependent variable (success or failure), we used a three-way loglinear analysis in order to assess statistical significance. For simplicity, one can think of a loglinear analysis as an ANOVA for categorical data: It determines the type of effect present across the three categorical variables.

The three-way loglinear analysis produced a final model that retained all effects. The likelihood ratio of this model was $\chi^2(0) = 0$, $p = 1$. This indicated that the highest-order interaction (model type × compression type × score) was significant, $\chi^2(1) = 5.37$, $p = .020$. To break down this effect, separate chi-square tests on the compression type were performed separately for each model. For Coherencer, there was a significant association between the compression type and success, $\chi^2(1) = 87.58$, $p < .000$.[11] This was also true for the top-*n* model, $\chi^2(1) = 74.96$, $p < .000$. Odds ratios indicated that the odds of success were 1.98 higher using co-occurrence probabilities for Coherencer, and 2.74 higher for top-*n*.

## 4. Discussion

The results support the hypothesis that Coherencer generates scenes that are more coherent than the top-*n* model and Thagard's connectionist model. They also support the notion that co-occurrence probabilities are a better association than holographic vectors and that better generative models (e.g., Coherencer) improve quality regardless of the association type used (e.g., co-occurrence probabilities, holographic vectors). In all cases, a "better result" is assumed to indicate what constraints better capture the required functionality that a human brain *could* implement. Again, we expect that humans can accurately infer higher-order associations with near perfect fidelity, so models that cannot even closely approximate this on a much simpler task are not likely candidates. However, the effect sizes in all cases were rather small, so the conclusions and speculations in this section are given tentatively.

On the basis of the results that show Coherencer's superiority to the top-*n* model, we can tentatively conclude that the mechanism for visual generative cognition in the human, visual imagination, cannot assume independence between its associated objects as per the top-*n* model. This means that models that assume such independence are not plausible candidates (e.g., naive Bayesian classifiers).

We can also conclude that both locally parallel approaches similar to Thagard's connectionist model as well as vector representations that use an addition relation and cosine similarity scores to approximate simple associations do not accurately model human beings in this task. In both cases, our serial algorithm outperformed these algorithms and people (we assume) are able to outperform them all.

Eliasmith (2013) specifically mentions that these vector models experience memory limitations when there are more than eight symbols in a 500-dimensional space and the number of symbols scales linearly with the number of dimensions. Our system uses thousands of symbols. There has been some effort to remedy these limitations using "chained" structures, however, and a more sophisticated holographic system might be able to overcome these limitations.

Similarly, Thagard's connectionist model has largely been surpassed by more state-of-the-art neural network approaches (e.g., deep networks). At present, we are unable to generalize the current findings to these new approaches. Our future work focuses on these models and how they might relate to the overall functionality of these networks and Coherencer's findings.

Before moving on to a discussion of related theoretical approaches, it is worth considering some limitations of the current model. At present, there are no higher-order relationships understood by the model (e.g., synonymy). So, Coherencer will often return synonyms of the query or other potentially problematic objects (e.g., internal parts of the query, more general descriptions of the query). This is certainly a limitation of the model in a general sense. However, at present, it is not entirely clear that simple associations (e.g., co-occurrence) can capture these properties, at least in the absence of sufficiently complex processes and computations. Thus, it is possible that, in the current context, these limitations are less a property of the model and more a property of the current domain. That is, humans likely use more than just simple associations to achieve coherence, especially when considering more complex properties like synonymy. Future work will need to resolve many of these issues.

### 4.1. Related theoretical approaches

Research on generative cognitive faculties traditionally focuses on determining constraints that will exclude combinations that humans do not generate from the total set of options that *could* be generated, much like Coherencer. We will describe three prominent models that select different constraints to achieve this goal: a top-down conceptual approach to constraints in word-pair concept combinations (the $C^3$ model; Costello & Keane, 2000), a modal approach to constraints through the use of simulations in perceptual symbol systems (Barsalou, 1999), and an amodal approach to constraint satisfaction from the literature on coherence (Thagard, 2002).

#### 4.1.1. The $C^3$ model

The work by Costello and Keane (2000) focuses on the generation and interpretation of modifier-head word pair combinations (e.g., "pickled cat") that are novel and often creative. They frame their work in terms of a "constraint theory of concept combination" (Costello & Keane, 1997). The theory proposes that concept combination is composed of two components—a "generative mechanism" and a set of three constraints: diagnosticity, plausibility, and informativeness. The generative mechanism builds *all* possible subsets of predicates (e.g., "pickled," "furry,") of the constituent (e.g., "cat") and related concepts, while the constraints discourage unlikely combinations and encourage highly likely combinations. They created the $C^3$ model to computationally test their approach.

The model finds a coherent (although the authors do not use the word) combination of predicates to *justify* a modifier-head word pair. Although this is a different task than that of Coherencer, $C^3$ has some similarities with Coherencer that makes the comparison valuable.

The $C^3$ model computationally implements the three constraints as follows. The diagnosticity constraint is represented by the joint relative frequency of the predicate and concept (i.e., $|C \cap P|/|C \cup P|$, where $C$ is the concept and $P$ is the predicate). Plausibility is represented by the average predicate overlap between stored instances of word pairs and the current set of predicates for a new word pair (i.e., $\frac{\sum_{A \in O} \|A\|/\|N\|}{|O|}$, where $O$ is the set of pairs whose predicates overlap with the new pair and $N$ is the new pair; note that $|A|$ gives the number of pairs while $\|A\|$ gives the number of predicates). Finally, a predicate set for a new word pair is informative if it is not strictly a subset of a stored instance of the prototype[12] of the concept (i.e., if $\mathrm{N} \not\subseteq \mathrm{H}$ or $\mathrm{N} \not\subseteq \mathrm{M}$, where $H$ is any stored instance of the head and $M$ is any stored instance of the modifier).

Coherencer's measure of co-occurrence is the conditional relative frequency of two objects in a scene (i.e., $|A \cap B|/|B|$ or $|A \cap B|/|A|$, where $\|$ gives the number of scenes containing the object). This is formally similar to the diagnosticity constraint despite the fact that Costello and Keane (2000) describe predicate overlap as the "co-occurrence" of predicates in a *different constraint*: the plausibility constraint. That is, they use the word *co-occurrence* in a formally different way. Even though diagnosticity is similar to our notion of co-occurrence, they are not the same. Unlike their measure of diagnosticity, *conditional* relative frequency is non-commutative, as we previously stated, whereas joint relative frequencies are commutative. We take this feature to be essential to cognition. At minimum, it assumes less about the world in that commutativity is more restrictive and, thus, less general.[13]

The other two constraints are problematic for our theoretical orientation to memory and storage. They assume that the stored representations are perfectly represented and accessible; otherwise they would not be usable for overlap comparison. This is inconsistent with the constructive memory hypothesis, which states that only traces of the complete memory are stored (Conway & Pleydell-Pearce, 2000; Rubin, Schrauf, & Greenberg, 2003; Schacter et al., 2012; Tulving & Watkins, 1973). Upon retrieval, the rest of the experience must be reconstructed from the traces. That is, these traces function as *lossy compressed representations*. This means that complete memories are *not* trivially accessible. Consequently, either Costello and Keane (2000) are modeling a different form of cognitive generation that occurs *after* the generative processes we are considering (a possibility we take to be unlikely) or they are excluding a fundamental component of the generative process.

Additionally, our new evaluation technique is not designed to evaluate output models that use the plausibility and informativeness constraints. Our technique exploits the lossyness of memory storage in order to quantitatively assess the mechanisms of cognitive generation. The $C^3$ model does not compress its memory. Interpreted in another way, our evaluation requires that the output of our model is perfectly plausible and uninformative, according to their constraints. However, we believe these constraints are cognitively *im*plausible, despite their name, for reasons mentioned above (e.g., they conflict with the constructive memory hypothesis).

The final point that we address concerning the work of Costello and Keane (2000) has to do with the control sequence of their $C^3$ model. The model first generates a set of partial predicate sets for a new word pair by selecting the most highly correlated predicates from the head and modifier concepts. Then, the model expands on these partial interpretations by adding new predicates from its knowledge base that have the highest plausibility while still being informative (i.e., not strictly subsets of stored representations). Due to the rather small size of their knowledge base (76 instances with 22 predicates on average), it is computationally tractable to compute the "best" predicate set for a new word pair. In the computational assessment, they discuss the model computes the top 10 predicate sets for each pair. A thresholding mechanism is also used to filter the output of their model.

Coherencer has some similarities and some differences to the $C^3$ model. It has a similar seeding process where top associations with the initial input are computed. It then expands this set using more loosely associated concepts much like the $C^3$ model.

However, its input database contains over 18,000 scenes and over 2,000 possible objects, so assessment of all possible combinations is intractable. Coherencer also uses a threshold, but it is more deeply integrated into the model: The threshold is used to determine when Coherencer should stop searching; it is not used to select which variation should be selected after they are all (or a large subset of variations) are generated.

We have several reasons to prefer Coherencer over $C^3$. $C^3$ implements an unrealistic theory of memory, in that it is uncompressed. It is unable to deal with large amounts of data, both because of memory and computational restrictions. Interestingly, the authors explicitly mention that an improvement on their model is Paul Thagard's parallel connectionist model—which Coherencer outperforms. Coherencer also better fits with what we know about the brain and its functioning, as will be discussed below.

#### *4.1.2. Modal approach: Perceptual symbol systems*

Barsalou's (1999) perceptual symbol systems (PSS) theory is an example of embodied theories of representation. Although this work is often viewed in direct contrast to more abstract approaches, like Coherencer, there is actually a great deal of overlap. In what follows, we briefly describe PSS and some of the commonalities between it and the Coherencer model.

The PSS theory initially challenges both the top-down and amodal, constraint-based approaches. Its approach integrates cognition intimately with perceptual processes that are explicitly modal and use analogy, both of which are properties that are related to the coherence problem, as we will argue below. According to PSS theory, the same neurocognitive systems involved in perception are also used for representation.

One of the central tenets of Barsalou's approach is his definition of a perceptual *symbol*, which can be described as follows. Unlike imagery or related conscious subjective experiences, a perceptual symbol is a record of the neural states that occur in perception. In this sense, it is a memory trace *of* perception and it corresponds to the compressed associations we discussed throughout the paper. These traces are stored via Hebbian strengthening or multisensory integration common to the hippocampus (another commonality with our theoretical approach, which we discuss below), and at no point do these traces ever completely transduce into an amodal symbol (Barsalou, 1999).

The traces, like the compressed representations, are only schematic (Barsalou, 1999). They do not encode a complete representation of the original experience, "only a very small subset that represents a coherent aspect of the state" (Barsalou, 1999). Consequently, citing Damasio (1989), Barsalou (1999) endorses something similar to the constructive view of memory and recall in order to account for the processes that work with these compressed, schematic, stored representations. According to Barsalou, there is a schematic structure for "mouse" that integrates (perhaps in an integrative region like the hippocampus) all the perceptual experiences of mice of various sensory modalities. Additionally, this schematic structure is *also* linked with a linguistic equivalent: "Thus, linguistic symbols index and control simulations to provide humans with … conceptual ability" (Barsalou, 1999).

According to Barsalou (1999), the schematic structure is a compressed representation (i.e., an index) of a holistic neural state (i.e., an experience) of a "mouse," for example. Linguistic symbols (e.g., the actual word "mouse") can then be coupled with these schematic indices such that they function as a second-order *reference* to the underlying holistic state(s).

For Barsalou and Damasio, the construction of these schematic indices in memory does not occur in the perceptual areas (i.e., the position of the holistic neural state), but in dedicated associative/integrative regions (e.g., the hippocampus and related areas). Damasio (1989) refers to the processing in this region as "amodal" because these areas do not "map sensory or motor activity in a way that preserves feature-based, topographic and topological relations of the external environment as they appear in psychological experience."[14] In contrast, Barsalou refers to these schematic indices as perceptual symbols.

Due to the separation between these, perhaps, *trans*modal (rather than "amodal") areas, it is possible to temporarily ignore the underlying holistic neural states in an effort to better understand how these associative convergence zones might function. The schematic index can then be used as an incomplete, but still useful, replacement for the holistic neural state exactly as it is done in the brain.

Although at first glance the labels used in Coherencer appear to be amodal and inconsistent with perceptual symbol's systems theory, as theoretical entities they correspond to the schematic indices of Barsalou and Damasio. This is particularly evident in Coherencer's role in the larger imagination engine SOILIE (Vertolli et al., 2014), in which Coherencer's label-based "symbols" are associated with hundreds of perceptual instances of the objects they represent and are not, by themselves, perceptual traces.

#### *4.1.3. Amodal approach: Thagard and coherence*

Thagard (2002) devotes an entire book to the coherence problem. He describes coherence as an optimization problem: the selection of the best combination of elements to optimize according to some set of criteria. Thagard takes these criteria to be a set of positive constraints (i.e., inclusion of one component increases the likelihood of inclusion of another component) and negative constraints (i.e., inclusion of one component decreases the likelihood of inclusion of another component). These constraints are optimized by maximizing the number of positive constraints in a collection and minimizing the negative constraints.[15]

This formalization of coherence is "amodal" in the same way that associative convergence zones for Damasio (1989) are "amodal." Mainly, it describes a type of functionality that could be applied to many different modalities. Interestingly, the commonalities do not stop there.

Thagard (2002) outlines a number of general classes of computational models that can resolve coherence problems. One class is the "incremental" algorithm. Incremental algorithms evaluate coherence at each time step relative to a single pool of selected elements. In Thagard's description, each element is evaluated only once when it is added to the pool.

Costello and Keane (2000), Barsalou (1999), and Coherencer use a similar approach with a few differences. First, the incremental algorithms described in Thagard (2002) build their initial pool one element at a time, whereas all the other models seed their initial pool with the strongest associations. Second, the space within the incremental algorithm's "working memory" can be of any size: It could literally contain the entire set of possible elements if that was what maximized coherence. Costello and Keane (2000) follow this approach when it comes to their predicates, while Coherencer has a finite limit on the size of its pool, which models the limits of human working memory. Barsalou (1999) does not specify. Third, Coherencer does not maximize coherence. It makes sure that it passes a certain threshold. Both other models use a maximization approach, like Thagard. Barsalou (1999), on the other hand, recognizes that constraints in the context can cause deviations from this global maximum, which aligns more closely with our position. Fourth, once an element has been selected by the incremental algorithm, it cannot be unselected. Costello and Keane (2000) do not describe their algorithm at this level of detail, but it seems like it is closer to Thagard's approach. Barsalou (1999) does not specify. Coherencer maintains backtracking capabilities for selected elements, but both Coherencer and the incremental algorithm cannot backtrack on rejected elements. The other models do not specify. We take Coherencer and all the other models to be variations of the incremental class of algorithms despite these differences: "Incremental" highlights the serial approach that is a defining feature of the models.

Incremental algorithms often lead to suboptimal solutions, much like Barsalou (1999) anticipated, by getting stuck on local optima. This is a direct consequence of the serial comparisons: These models are limited to a local perspective, so they often reject elements that are optimal at a global level. Although backtracking for these rejected elements has been implemented as a fix for serial approaches, Thagard suggests that this is worse than some alternatives.

Thagard (2002) proposes a connectionist model in response to this problem. These models examine solutions in parallel, which decreases the probability of getting stuck on suboptimal solutions. Consequently, we expected connectionist models to outperform incremental algorithms like Coherencer and $C^3$ on coherence-related problems. Although we found that they did not in our particular instantiation of the problem, Thagard has implemented a number of connectionist models with success (e.g., Eliasmith & Thagard, 1997; Thagard, 1989, 1991, 1992a,b, 2002; Thagard, Holyoak, Nelson, & Gochfeld, 1990). However, he does leave one caveat for the weaker class of incremental algorithms. Thagard points out that they have one important contribution: They can offer valuable insights into human cognition, which is known to perform suboptimally in many domains, including coherence.

We add to this position the idea that, like Barsalou (1999) recognizes, suboptimality is often indicative of a more extensive system of constraints. The essence of being extremely sensitive to local conditions *means* that one *cannot* be globally optimal. This is the fundamental idea behind the no free lunch theorem for theories of optimization in computer science (Wolpert & Macready, 1997). With increasing specificity and precision, one loses general applicability and vice versa.

Recall that our theory is that Coherencer's serial approach to inferring coherence through higher-order dependency relations provides some of the necessary requirements for a generative mechanism of imagination. We argue that these same features give associative areas in the brain, specifically the hippocampus, the necessary requirements to preserve semantic constraints while generating idiosyncratic instantiations of those constraints. This theory and the Coherencer model have a lot in common with broader theoretical, computational, and neuroscientific literatures. We will address some of these points of overlap before discussing some predictions of the model.

### 4.2. Coherencer and the brain

Some of the research discussed above suggests an involvement of the hippocampus in generative mechanisms in the brain. We argue that the conversion process from semantic memory to episodic memory to simulated experience (i.e., abstract to modal internal representations), in particular, is mediated by this neural structure.[16] The argument integrates many of the threads we have discussed above. We will begin by addressing some of the earlier theories of the hippocampus and then move to more contemporary theories, especially scene construction theory (SCT).

Research on the distinction between semantic dementia and Alzheimer's disease suggests that the hippocampus is much more strongly associated with autobiographical details (i.e., episodic memories) than semantic memory proper (Chan et al., 2001).

More recent research into remote memory extends this basic insight. Remote memory is one of two types of long-term memory (the other is synaptic consolidation), which encompasses memories that last more than 24 hours (Dudai, 2004b). The two types are differentiated by the length of time they take to consolidate and stabilize in the brain on average, as well as the underlying mechanisms. Remote memory takes days to decades to stabilize and uses the hippocampus. In particular, remote memory affects systems-level information traces distributed across the brain (Dudai, 2004b; Frankland & Bontempi, 2005). There are three main theories of remote memory: the cognitive map theory, the standard consolidation theory, and the multiple trace theory.

Cognitive map theory claims that the hippocampus creates the context of episodic memories through allocentric spatial representations (Burgess, Maguire, & O'Keefe, 2002; Shimamura, Squire, & Shacter, 2002). This *always* occurs for both recent and remote memories, and it does not occur for semantic memories, as they are context-free.

The standard consolidation theory posits that the hippocampus and related areas are used to consolidate memory traces to the neocortex and other extrahippocampal structures (Dudai, 2004b; Frankland & Bontempi, 2005). Memory consolidation is defined as the progressive process by which long-term memories are stabilized post-acquisition (Dudai, 2004a). In this view, the hippocampus is no longer required after consolidation (in direct opposition to cognitive map theory). No distinction is made between semantic and episodic memories, so it is assumed that both undergo the same process, which also contradicts cognitive map theory.

The "multiple trace theory" claims that the hippocampus's primary function is the re-experiencing of past, autobiographical events regardless of their age (Moscovitch, Rosenbaum, et al., 2005; Moscovitch, Westmacott, et al., 2005; Moscovitch, Nadel, Winocur, Gilboa, & Rosenbaum, 2006). In this view, each trace acts as an index to neocortical regions (e.g., temporal cortex for object details) where the details of the relevant event are represented and stored. Memory retrieval is viewed as a form of re-encoding that results in older memories having a wider distribution of traces throughout the hippocampal complex. Semantic memory requires hippocampal processing early on, after which the cortex dominates. Spatial, cognitive maps (i.e., spatial contexts) can be encoded in the cortex in a schematic form of semantic memory, in direct contrast to cognitive map theory's position.

Multiple trace theory both integrates and expands the consolidation theory and cognitive map theory. It also aligns closely with the constraint-based views already discussed. It endorses the constructive memory hypothesis, emphasizes the re-experiential role of the hippocampus, and highlights the importance of abstractive mechanisms for both schemas and the consolidation of semantic memories.

With this in mind, we take generative imagination to be a reversal of the memory storage process. Thus, instead of going from perception to episodic then semantic memory, generative processes go from the highly abstract semantic memory to the schematic traces of episodic memory and then the internal simulations of imaginary simulation (i.e., the final, re-experiencing of some imagined event). Given how central the hippocampus is to the memory storage process, our expectation is that it should also be central to the generative processes of imagination. The contemporary view of hippocampal function called "SCT" begins to specify what these hippocampal processes might look like. In what follows, we will explore this theory of hippocampal function in detail in order to indirectly explore how cognitive generation might occur in the brain.

Like multiple trace theory, SCT proposes that the hippocampus plays a role in the creation of a "coherent spatial context" in which details of episodic memories can be "martialed, bound and played" (Hassabis & Maguire, 2007; Maguire & Mullally, 2013). Unlike multiple trace theory, SCT explicitly recognizes the role of the hippocampus in imagined future experiences. In SCT, the hippocampus operates as an anticipatory structure much like Barsalou's (1999) "conceptual system," rather than just a memory (i.e., "recording") system. In this role, it provides a "cohesive spatial framework" for "what is outside the direct field of experience" that is grounded in what is processed in one's direct experiential field. One example is being aware of the back of the chair that one is sitting on, when not touching or visually observing it.

The hippocampus has been implicated in cognitive functions as diverse as spatial navigation, future-thinking, and imagination, in addition to its traditional role in memory (Addis, Pan, Vu, Laiser, & Schacter, 2009; Botzung, Denkova, & Manning, 2008; Maguire & Mullally, 2013; Szpunar, Watson, & McDermott, 2007). It is not that the hippocampus is the only area of the brain responsible for these cognitive functions, but that it provides an essential ingredient to all of them: the ability to create coherent, spatial scenes.

Although there has been some debate in the literature on this topic (Hassabis, Kumaran, & Maguire, 2007; Maguire & Hassabis, 2011; Squire et al., 2010; Squire, McDuff, & Frascino, 2011), SCT is also supportive of more recent research on a phenomenon called "boundary extension" (Chadwick, Mullally, & Maguire, 2013; Maguire & Mullally, 2013).[17] Boundary extension is the filling out of an image or scene beyond the edges of that scene. Boundary extension only occurs when viewing a complete scene (as opposed to individual objects). People with damage to their hippocampus do not experience this effect, suggesting that the hippocampus is associated with one of the distinctions between an independent object (e.g., a rock on a blank background or no background at all) and an object as part of a scene.

The SCT is not a model that generates quantitative predictions of the information processing of the underlying neuroscience. An information processing account would allow for the prediction of more detailed and specific hypotheses (Kumaran & Maguire, 2009).

#### *4.2.1. Memory and cognitive generation in the hippocampus*

As discussed, generative processes are necessary for various cognitive functions (e.g., imagination, future thinking, spatial navigation). Thus, we, like SCT, view many of the neurocognitive processes used in generation on a continuum from normal memory retrieval processes to the fully generative constructions of imagination. Hassabis and Maguire (2007) draw on three lines of evidence for this position: the constructive memory hypothesis; a similar network of brain areas used in the recall of real, recall of fake, and imagination of new experiences; and the top-down effects of hippocampal processing on other brain areas. We address and expand these points through comparisons with our proposed framework.

The constructive memory hypothesis states that the recollection of past experiences is a generative reconstruction on the basis of traces in long-term memory (Conway & Pleydell-Pearce, 2000; Rubin et al., 2003; Schacter et al., 2012). This is in direct contrast to theories that propose memories are stored and retrieved in their entirety (e.g., see Brewer & Dupree, 1983). Hassabis and Maguire (2007) argue that the constructive view is more plausible on computational grounds (e.g., due to storage constraints and the needs of generalization and abstraction) as well as behavioral (e.g., it accounts for well-known memory errors).

For the next point, Hassabis and Maguire (2007) show a consistent pattern of neural activation across imagination and scene construction phenomena. In a conjunction analysis comparing the recall of real memories, fictitious memories created the week prior, and new fictitious scenes imagined for the first time, the hippocampus, parahippocampal gyrus, retrosplenial cortices, posterior parietal cortices, ventromedial prefrontal cortex, and medial temporal cortices are all consistently active (Hassabis et al., 2007). This was found in direct contrast to control conditions that required no scene construction (i.e., real and imagined simple objects). In a position similar to Slotnick et al. (2012), Hassabis and Maguire (2007) argue that it is this same network that has been implicated in all the various processes closely associated with episodic memory (e.g., navigation, spatial reasoning, imagination). Many of the same brain regions are active across the scene construction and imagery research as well.

Hassabis and Maguire (2007) also show that the leading alternatives for hippocampal function—the subjective sense of time, a sense of selfhood, and autonoesis (i.e., mental time travel and related phenomena)—are more closely associated with other brain areas. SCT directly contrasts with claims that position them as the primary function of the hippocampus (for one of these competing views, see Tulving, 2002).

In the previous two sections, we highlighted that what is unique to generative imagination is that it goes beyond what was originally stored in memory as per the constructive memory hypothesis. Consistent with SCT, generation is more broadly applicable than memory, in that it does not have to index a specific event in one's past (i.e., episodic memory retrieval). Our theory of imagination expands on SCT by incorporating Thagard's notion of optimization-based coherence. Thus, memory construction is algorithmically described as a form of optimization—or, more broadly, satisficing (Simon, 1956)—of a combinatorial arrangement using semantic constraints. This optimization is directly anticipatory. That is, it answers the question, given objects $x$, $y$, and $z$, what other objects are likely to be present. Since this anticipatory function can be broadly applied, we described it as an amodal form of optimization in the previous section.

Although SCT does not explicitly use co-occurrence in their description of the hippocampus, they do use spatial properties such as proximity. As we stated in the introduction, co-occurrence is a highly compressed (i.e., general or abstract) representation of many of these spatial properties: Every spatial relation implies that the objects possessing that relation co-occur. Co-occurrence captures spatial as well as other associative properties. Consequently, it is consistent with SCT.

Finally, Coherencer and SCT have a similar processing sequence that is consistent with both perceptual symbol systems and mental imagery. For SCT, the construction of anticipated extensions of the current scene in the hippocampus is propagated back down the network to change the perceptual neural firing in the visual cortex towards the extended view (Chadwick et al., 2013). This then propagates back to the hippocampus in a feedback loop.

All of the models discussed so far do something similar, if at a very abstract level. An anticipated combination of objects (corresponding to a pattern of neuronal firing in the visual cortex) is retrieved from a bank of possibilities in memory. Their associations are assessed for coherence (corresponding to hippocampal processing). If it is not coherent enough, one of the selected objects is swapped for a different possible object (i.e., a neuronal pattern in the visual cortex).

An illustrative example is the task of moving through a dark room. If one started in the bedroom, it would be highly unlikely to anticipate running into the kitchen table. These are not the most highly associated objects with a bedroom. Consequently, most people (with similar historical and cultural housing contexts) would probably anticipate a hallway, carpet, walls, and maybe a distant railing for the stairs in their representation of the space. If they step on part of a child's toy, then they would likely update the space with the as yet unexperienced parts of the toy, further toys, or related objects.

The experience of the toy adds a new element to the context that further constrains anticipatory processing.

*4.2.2. Perceptual symbols*

Another parallel between SCT and our framework has to do with the relationship between hippocampal processing and systems used by perception. Kumaran and Maguire (2009) argue that memory and perception are not discrete neural modules. The hippocampus acts on neural firing in the visual cortex to create the experience of anticipatory extension (Chadwick et al., 2013). Therefore, perception is motivated by neural processing and memory as well as by environmental input.

Research by Howard, Kumaran, Ólafsdóttir, and Spiers (2011) extended Kumaran and Maguire's work by differentiating the CA1 and CA3 subregions of the hippocampus. According to Howard et al. (2011), the CA1 receives sensory input from the entorhinal cortex and input from memory through the CA3 subregion. Perception informs memory and memory informs perception as Barsalou (1999) argued. The CA1 operates as an associative match-mismatch detector that responds when a pattern is recognized, the rest of the pattern is anticipated, and then the anticipation is violated (e.g., A-B-C-D is shown then A-B-D-C).

The seeding process proposed by Coherencer and the other models discussed is close in functionality to the CA3 associative subregion in that both are the interface with memory retrieval and anticipatory processes. Although memory retrieval and anticipatory processes are distinct, they intimately inform one another: Coherencer cannot return anything without the initial structure (i.e., the most frequently co-occurring objects for a given query) that is returned from memory. The thresholding mechanism, which mirrors the CA1 by detecting whether or not the set is coherent, then closely informs when new objects need to be selected from memory. Thus, the models provide a much more detailed description of the neural process at a functional scope.

*4.2.3. Coherence or spatial navigation*

Mullally, Intraub, and Maguire (2012) found that patients with damage to the hippocampus were unable to experience boundary extension. However, they were perfectly able to describe an appropriate context for the scene or additional objects that would likely be present in the scene if they were to mentally "take a step back." Outside of the boundary extension itself, they were also unable to determine where objects would be spatially located relative to one another: to spatially integrate the scene. The subjective quality of their imagined experience was also more limited, as given by lower self-report scores. Superficially, this appears to be evidence of separation between Coherencer and SCT. However, we argue that this is not the case.

First, when the authors are talking about spatial integration, they mean a scene that possesses a spatial arrangement of the objects that is appropriate to the real world (Maguire, personal communication). If an animal mouse is on a mouse pad on a desk with a computer (i.e., there is an explicit, realistic, and describable relationship between the mouse and every other object in the scene), it is spatially integrated. Second, the concept of spatial integration is related but distinct from Coherencer's concept of coherence. Spatial properties assume co-occurrence. What is unclear is whether the "additional objects" that the patients could generate were coherent. Given our basic model of the process, the association between spatial integration and coherence through co-occurrence, and the fact that the patients could not spatially integrate, we predict that the additional objects were incoherent (or had a very low coherence score) with one another. Nevertheless, Mullally, Intraub, and Maguire claim that there is some association.

As we will discuss in detail in one of the final sections on the coherence of spatial integration, it is our suspicion that the collection of these objects as a whole would be incoherent: The associations returned are actually the most strongly associated objects initially queried by Coherencer. They would appear qualitatively associated in that they combine well with the original query, while still being of low coherence overall. The set of responses given by patients in the assessment of coherence was not reported by Mullally et al. (2012). Thus, one can only speculate at present.

### *4.3. Predictions and implications*

As a cognitive model, Coherencer makes predictions concerning (a) serial processing, (b) the distinction between memory retrieval and generation, and (c) spatial integration. We discuss each of these predictions in turn.

#### *4.3.1. Serial structure*

Coherencer and similar incremental models predict that objects get placed into and taken out of the object set during the coherence process in a serial manner. That is, the image goes through several drafts before the final version is arrived at. An empirical test of this might be challenging because we are typically conscious of only the final image. From an imaging perspective, we do not yet know what subpatterns of activation in the brain are corresponding to what objects, making it difficult to measure what objects are placed and then removed from intermediate images.

However, it should be possible to map out, for a given individual, the most likely associations with certain objects. Input contexts could then be designed that force certain incoherent associations, which should appear, residually, in the output. For example, if "mouse" is queried and it is most closely associated with "cheese" and "computer," while "computer" is most closely associated with "screen" and "bank," one might expect to get an output that is informed by all of these associations even if "mouse," "cheese," and "bank" are the only ones actually selected. For example, the resulting output might be an animal "mouse" eating "cheese" in a "bank" office that has computers. In this case, "bank" would occur as a consequence of its strong association to "computer," even though it might have a weaker association to "mouse" than other labels.

Another important aspect of serial processes is that harder problems will take longer than easier problems. When a problem is harder, more incorrect selections will be made before the correct answer is selected and every additional incorrect answer in a serial process requires more time. Thus, if generative processes in imagination are serial, then queries with more incoherent associations (e.g., a greater number of different contexts that overlap) should take longer to process or at least take proportionally longer relative to the number of incorrect selections one would expect to occur.[18] This is empirically testable.

#### 4.3.2. Memory retrieval

In the hippocampus section, we briefly touched on some of the differences between memory retrieval and generative imagination. Our suggestion was that the difference between the two of them is a matter of degree. Both generate information lost during the storage process using associative constraints from semantic memory. But normal memory retrieval includes local (as opposed to global or general) constraints on the scene that are temporally indexed (i.e., a particular place in time in one's life; for a similar position, see Tulving, 2002). Generative imagination is not required to use these constraints in the general case, and so generative cognition can be more broadly applied across tasks (e.g., imagining fictional entities, which have no temporal index). This should be directly testable by comparing subjects with damage to regions more directly associated with recall of the past (e.g., frontoparietal regions) and those with damage to the hippocampus alone (Hassabis & Maguire, 2007). We therefore predict that hippocampal damage will have broader effects corresponding to its more global, associative mechanisms (e.g., general coherence), whereas damage to regions that perform recollection of the past should show more local deficits related to constraints that are required for memory retrieval (e.g., temporal indexing).

The basic theory that we are proposing is that coherence is derived from semantic associations in memory. These semantic associations are originally abstracted through hierarchical transitions of perceptual processes (e.g., through the functions of the visual system) and then higher association areas in the brain (e.g., the hippocampus). In an instance of imagination, these semantic associations are converted back to more concrete representations that are common to episodic memory and more concrete imagery processes. In this view, semantic and episodic memory retrieval, and mental imagery and simulation exist on a continuum of abstraction with semantic memory on the abstract end, episodic memory more concrete, and mental imagery the most concrete. The more concrete the representation, the greater the contribution of the corresponding modal processes (e.g., visual perception) as opposed to semantic structures distributed among more disparate cortices (e.g., temporal lobe, frontal lobes). Generative imagination, then, captures in large part this conversion from conceptual representation to scene descriptions to depictive processes of mental imagery.

#### 4.3.3. The coherence of spatial integration

Another test of Coherencer's extension of SCT requires the verbal data, like those collected in Mullally et al. (2012)—of the responses of the patients and controls when asked what objects would be present if they took a step back. The sets could then be input into the model to assess their level of coherence. Coherencer predicts that their coherence would be low. Patients with hippocampal damage will produce incoherent images in their imagination. If the results show that patients can generate combinations of objects with high coherence, it would suggest that Coherencer needs to be analyzed at a finer scope. Some of its processes would be shifted down the line of processing closer to memory, while others (e.g., anticipatory functions) would remain within the functionality of the hippocampus proper.

These differences are already present in the basic structure of Coherencer. When given a query, the first step performed by the model is to seed its buffer with the objects that most frequently co-occur with the query. The model assumes that given a query, human memory is able to retrieve these high-frequency objects, corresponding to the initial, top-*n* selection in the model. Consequently, the top-*n* selection process is already very proximate to memory while later anticipatory coherence processes are more distant. What one would expect on the basis of this model is that patients with hippocampal damage would use the objects recalled in this seeding process to form their response.[19] All of these objects would be locally coherent with the query (e.g., "mouse"). However, in as much as the query could occur in multiple different contexts, the cluster of objects given will have a low coherence: The patients will be relying on only the initial seeding as a consequence of skipping the contextual refinement that Coherencer models.

### 4.4. Conclusion

This work has focused on a number of contributions to the understanding of cognition, computation, and neuroscience. The overarching goal was to discuss in detail the notion and mechanisms of coherence in terms of human visual imagination. Our results allow us to tentatively conclude that Coherencer generates scenes that are more coherent than models that assume independence between co-occurring items (e.g., top-*n* model) as well as approaches that converge on a solution via parallel processes (e.g., Thagard's connectionist model). We can also tentatively conclude that co-occurrence probabilities are a better association than holographic vectors—at least for the current evaluation—and that better generative models (e.g., Coherencer) improve quality across association types (e.g., co-occurrence probabilities, holographic vectors).

Theoretically, we have shown that there is a great deal of overlap between the approach used in Coherencer and other models of cognitive generation. In particular, we showed that Coherencer is both very similar *and* more cognitively plausible than the $C^3$ model created by Costello and Keane (2000). We also showed that Coherencer is consistent with *both* the modal schematic indices of perceptual symbol system theory (Barsalou, 1999) and the amodal contextual constraints espoused of Thagard's (2002) formal model of coherence. Finally, we demonstrated that Coherencer is also consistent with contemporary research on the hippocampus: a neural system that is implicated in both imagination and cognitive generation, more broadly.

In sum, the Coherencer model does a great deal of work illustrating the types of systems that could viably be used to achieve coherence in the visual imagination. At present, this work functions entirely at the computational and algorithmic levels of Marr and Poggio's (1976) trilevel hypothesis. Future work will focus on instantiating these insights at an implementational level—the underlying neural instantiation. From there, it will then be possible to make subtler differentiations in the underlying coherence mechanisms in the visual imagination.

## Notes

1. We give a formal definition of the coherence problem in Vertolli and Davies (2014). It has broad application in many areas of human cognition, much like imagination, but the two topics do not perfectly overlap.
2. The constructive memory hypothesis suggests that this is probably true of episodic memory as well (Conway & Pleydell-Pearce, 2000).
3. By selecting the label with the highest conditional probability, which is what the top-*n* model does, we are effectively implementing a naïve Bayesian classifier. That is, it assumes that the four selected labels are conditionally independent. It then selects on the basis of conditional probability with the query.
4. It is worth noting that the human mind is flexible enough to create narratives that make sense of hybrid combinations (e.g., a doctor administering an injection to an injured surfer after a 911 call requiring police support). However, what is important here is that the original experiences (i.e., the overall context of the corresponding observer) do not provide sufficient information to warrant the creation of these hybrid contexts. They are in some sense incoherent or fantastical relative to the limited range of images SOILIE has. The life history of a given observer determines what is considered coherent in the total space of possible combinations, and there is no agent-independent measure of whether a generated image is coherent. It depends on the agent's memory.
5. Note that although this link is usually bidirectional, the weight for each direction will be different: The probability of label A given B is rarely equal to B given A. Conditional probability is non-commutative as we stated above.
6. $P_{mean}$ was found to be 0.097 for our dataset.
7. All formulas are vector implementations of those described by Thagard (2002). We chose to use row vectors instead of column vectors as this more closely mirrors Coherencer's implementation.
8. $\theta$ was selected based on preliminary analyses of the model.
9. Higher dimensional vectors better approximate orthogonality between the labels (Cai, Fan, & Jiang, 2013). However, values near 1,000–2,000 have been shown to have lower error rates in memory tasks. It is for this reason that we selected a dimensionality of 1,000 (Rutledge-Taylor, Kelly, West, & Pyke, 2014).
10. Although the scores are low overall, it is worth keeping in mind that the models are probably creating many coherent label sets that are not found in the test set. This pattern continues for all subsequent tests.
11. We chose not to include the chi-square tables as they can easily be inferred from Figure 4.

12. The authors never fully define what a prototype is. However, they qualify that the prototypes were either automatically generated from instances in the knowledge base or hand coded.
13. Recall that parallel research on co-occurrence in machine learning suggests that non-commutative approaches are more realistic (e.g., almost all weddings have flowers but most flowers are not in weddings) and most models *assume* commutativity for mathematical or algorithmic convenience (Huang et al., 2012; Zhang & Zhou, 2013).
14. Damasio (1989) explicitly uses the word "amodal" in his discussion of the convergence zones. We expect that Barsalou (1999) avoids this language in order to maintain a clean distinction between Damasio's deeply perceptual amodality and the traditional, transduced symbol approach. We reintroduce Damasio's usage of the term in order to more subtly differentiate our position from previous research, as will become clear.
15. Note that in our framework, negative constraints are weak associations (i.e., two labels inhibit one another if their co-occurrence is zero). We do not directly deal with negative relationships in our dataset (e.g., that water could never occur as a fluid in space in the absence of a strong heat source). However, we expect that both of these properties will play an important role in future models.
16. However, we leave open whether simulated experience can occur in the absence of this abstract to modal conversion.
17. It is worth noting that Squire is the main opposition to SCT and the main neuroscience reference for both the associative regions of Barsalou (1999) and critique of a memory-only view of the hippocampus by Slotnick, Thompson, & Kosslyn (2012).
18. Parallel systems might also take longer to converge when there are more incoherent solutions, but the relation would likely be more complex and it probably would not be linear.
19. Sheldon, Romero, and Moscovitch (2013) found results that are similar to what we are proposing. Their study showed that patients with medial temporal lobe amnesia, which includes the hippocampus, were more likely to produce highly similar words to the query than controls in a free association task with words that had low semantic content (i.e., required the hippocampus). We suspect this is also a top-*n* like functionality at work in a lexical rather than visual context. However, preliminary exploration of the dataset has been challenging due to minimal overlap between the lexical categories and Coherencer's training set.

## References

Addis, D. R., Pan, L., Vu, M.-A., Laiser, N., & Schacter, D. L. (2009). Constructive episodic simulation of the future and the past: Distinct subsystems of a core brain network mediate imagining and remembering. *Neuropsychologia*, *47* (11), 2222–2238.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, *61* (3), 183.

Barlow, H. B. (1961). Possible principles underlying the transformations of sensory messages. In W. Rosenblith (Eds.), *Sensory communication* (ch. 13, pp. 217–234). Cambridge, MA: MIT Press.

Barsalou, L. W. (1999). Perceptual symbol system. *Behavioral and Brain Science*, *22*, 577–660.

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, *13* (February), 281–305.

Botzung, A., Denkova, E., & Manning, L. (2008). Experiencing past and future personal events: Functional neuroimaging evidence on the neural bases of mental time travel. *Brain and Cognition*, *66* (2), 202–212.

Breault, V., Ouellet, S., Somers, S., & Davies, J. (2013). Soilie: A computational model of 2D visual imagination. In R. West & T. Stewart (Eds.), *12th international conference on cognitive modeling* (vol. *196*, pp. 95–100). Ottawa, ON: Carleton University.

Brewer, W. F., & Dupree, D. A. (1983). Use of plan schemata in the recall and recognition of goal-directed actions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9* (1), 117.

Burgess, N., Maguire, E. A., & O'Keefe, J. (2002). The human hippocampus and spatial and episodic memory. *Neuron*, *35* (4), 625–641.

Byrne, M. D. (2013). How many times should a stochastic model be run? An approach based on confidence intervals. In R. West & T. Stewart (Ed.), *12th international conference on cognitive modeling* (vol. *196*, pp. 143–196). Ottawa, ON: Carleton University.

Cai, T., Fan, J., & Jiang, T. (2013). Distributions of angles in random packing on spheres. *The Journal of Machine Learning Research*, *14* (1), 1837–1864.

Chadwick, M. J., Mullally, S. L., & Maguire, E. A. (2013). The hippocampus extrapolates beyond the view in scenes: an FMRI study of boundary extension. *Cortex*, *49* (8), 2067–2079.

Chan, D., Fox, N. C., Scahill, R. I., Crum, W. R., Whitwell, J. L., Leschziner, G., Rossor, A. M., Stevens, J. M., Cipolotti, L., & Rossor, M. N. (2001). Patterns of temporal lobe atrophy in semantic dementia and Alzheimer's disease. *Annals of Neurology*, *49* (4), 433–442.

Cockbain, J., Vertolli, M. O., & Davies, J. (2014). Creative imagination is stable across technological media: The spore creature creator versus pencil and paper. *The Journal of Creative Behavior*, *48* (1), 13–24.

Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological Review*, *107* (2), 261.

Costello, F. J., & Keane, M. T. (1997). Polysemy in conceptual combination: Testing the constraint theory of combination. In M. G. Shafto & P. Langley (Ed.), *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society* (pp. 137–142). Hillsdale, NJ: Lawrence Erlbaum Associates.

Costello, F. J., & Keane, M. T. (2000). Efficient creativity: Constraint-guided conceptual combination. *Cognitive Science*, *24* (2), 299–349.

Cowan, N. (2001). Metatheory of storage capacity limits. *Behavioral and Brain Sciences*, *24* (01), 154–176.

Damasio, A. R. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition*, *33* (1), 25–62.

Davies, J., Atance, C., & Martin-Ordas, G. (2011). A framework and open questions on imagination in adults and children. *Imagination, Cognition and Personality*, *31* (1), 143–157.

Dudai, Y. (2004a). *Memory from a to z: Keywords, concepts, and beyond*. Oxford, UK: Oxford University Press.

Dudai, Y. (2004b). The neurobiology of consolidations, or, how stable is the engram? *Annual Review of Psychology*, *55*, 51–86.

Edin, F., Klingberg, T., Johansson, P., McNab, F., Tegnér, J., & Compte, A. (2009). Mechanism for top-down control of working memory capacity. *Proceedings of the National Academy of Sciences*, *106* (16), 6802–6807.

Eliasmith, C. (2013). *How to build a brain: A neural architecture for biological cognition*. New York: Oxford University Press.

Eliasmith, C., & Thagard, P. (1997). Waves, particles, and explanatory coherence. *The British Journal for the Philosophy of Science*, *48* (1), 1–19.

Eliasmith, C., Stewart, T. C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., & Rasmussen, D. (2012). A large-scale model of the functioning brain. *Science*, *338* (6111), 1202–1205.

Frankland, P. W., & Bontempi, B. (2005). The organization of recent and remote memories. *Nature Reviews Neuroscience*, *6* (2), 119–130.

Hassabis, D., & Maguire, E. A. (2007). Deconstructing episodic memory with construction. *Trends in Cognitive Sciences*, *11* (7), 299–306.

Hassabis, D., Kumaran, D., & Maguire, E. A. (2007). Using imagination to understand the neural basis of episodic memory. *The Journal of Neuroscience*, *27* (52), 14365–14374.

Howard, L. R., Kumaran, D., Ólafsdóttir, H. F., & Spiers, H. J. (2011). Double dissociation between hippocampal and parahippocampal responses to object–background context and scene novelty. *The Journal of Neuroscience*, *31* (14), 5253–5261.

Huang, S.-J., Yu, Y., & Zhou, Z.-H. (2012). Multi-label hypothesis reuse. In Q. Yang, D. Agarwal & J. Pei (Eds.), *Proceedings of the 18th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 525–533). New York: ACM Press.

Jones, M. N., & Mewhort, D. J. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, *114* (1), 1.

Kosslyn, S. M. (1996). *Image and brain: The resolution of the imagery debate*. Cambridge, MA: MIT Press.

Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The case for mental imagery*. New York: Oxford University Press.

Kumaran, D., & Maguire, E. A. (2009). Novelty signals: A window into hippocampal information processing. *Trends in Cognitive Sciences*, *13* (2), 47–54.

Maguire, E. A., & Hassabis, D. (2011). Role of the hippocampus in imagination and future thinking. *Proceedings of the National Academy of Sciences*, *108* (11), E39–E39.

Maguire, E. A., & Mullally, S. L. (2013). The hippocampus: A manifesto for change. *Journal of Experimental Psychology: General*,*142* (4), 1180.

Markman, K. D., Klein, W. M., & Suhr, J. A. (2012). *Handbook of imagination and mental simulation*. New York: Psychology Press.

Marr, D., & Poggio, T. (1976). *From understanding computation to understanding neural circuitry*. Cambridge, MA: MIT.

Moscovitch, M., Rosenbaum, R. S., Gilboa, A., Addis, D. R., Westmacott, R., Grady, C., McAndrews, M. P., Levine, B., Black, S., Winocur, G., & Nadel, L. (2005). Functional neuroanatomy of remote episodic, semantic and spatial memory: a unified account based on multiple trace theory. *Journal of Anatomy*, *207* (1), 35–66.

Moscovitch, M., Westmacott, R., Gilboa, A., Addis, D. R., Rosenbaum, R. S., Viskontas, I., Priselac, S., Svoboda, E., Ziegler, M., Black, S., Gao, F., Grady, C., Freedman, M., Kohler, S., Leach, L., Levine, B., McAndrews, M. P., Nadel, L., Proulx, G., Richards, B., Ryan, L., & Stokes, K. (2005). Hippocampal complex contribution to retention and retrieval of recent and remote episodic and semantic memories: Evidence from behavioral and neuroimaging studies of healthy and brain-damaged people. In N. Ohta, C. M. MacLeod, & B. Uttl (Eds.), *Dynamic cognitive processes* (pp. 333–380). Tokyo: Springer-Verlag.

Moscovitch, M., Nadel, L., Winocur, G., Gilboa, A., & Rosenbaum, R. S. (2006). The cognitive neuroscience of remote episodic, semantic and spatial memory. *Current Opinion in Neurobiology*, *16* (2), 179–190.

Mullally, S. L., Intraub, H., & Maguire, E. A. (2012). Attenuated boundary extension produces a paradoxical memory advantage in amnesic patients. *Current Biology*, *22* (4), 261–268.

Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences*, *25* (02), 157–182.

Rubin, D. C., Schrauf, R. W., & Greenberg, D. L. (2003). Belief and recollection of autobiographical memories. *Memory & Cognition*, *31* (6), 887–901.

Rutledge-Taylor, M. F., Kelly, M. A., West, R. L., & Pyke, A. A. (2014). Dynamically structured holographic memory. *Biologically Inspired Cognitive Architectures*, *9*, 9–32.

Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., & Szpunar, K. K. (2012). The future of memory: Remembering, imagining, and the brain. *Neuron*, *76* (4), 677–694.

Sheldon, S., Romero, K., & Moscovitch, M. (2013). Medial temporal lobe amnesia impairs performance on a free association task. *Hippocampus*, *23* (5), 405–412.

Shimamura, A. P., Squire, L., & Shacter, D. (2002). Relational binding theory and the role of consolidation in memory retrieval. In L. R. Squire & D. L. Schacter (Eds.), *Neuropsychology of memory* (3rd ed., pp. 61–72). New York: Guilford.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63* (2), 129.

Slotnick, S. D., Thompson, W. L., & Kosslyn, S. M. (2012). Visual memory and visual mental imagery recruit common control and sensory regions of the brain. *Cognitive Neuroscience*, *3* (1), 14–20.

Squire, L. R., van der Horst, A. S., McDuff, S. G., Frascino, J. C., Hopkins, R. O., & Mauldin, K. N. (2010). Role of the hippocampus in remembering the past and imagining the future. *Proceedings of the National Academy of Sciences*, *107* (44), 19044–19048.

Squire, L. R., McDuff, S. G., & Frascino, J. C. (2011). Reply to Maguire and Hassabis: Autobiographical memory and future imagining. *Proceedings of the National Academy of Sciences*, *108* (11), E40–E40.

Szpunar, K. K., Watson, J. M., & McDermott, K. B. (2007). Neural substrates of envisioning the future. *Proceedings of the National Academy of Sciences*, *104* (2), 642–647.

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences*, *12*(3), 435–467.

Thagard, P. (1991). The dinosaur debate: Explanatory coherence and the problem of competing hypotheses. In J. Pollock & R. Cummins (Eds.), *Philosophy and AI: Essays at the interface* (pp. 279–300). Cambridge, MA: MIT Press.

Thagard, P. (1992a). Adversarial problem solving: Modeling an opponent using explanatory coherence. *Cognitive Science*, *16* (1), 123–149.

Thagard, P. (1992b). *Conceptual revolutions*. Princeton, NJ: Princeton University Press.

Thagard, P. (2002). *Coherence in thought and action*. Cambridge, MA: MIT Press.

Thagard, P., & Stewart, T. C. (2011). The aha! experience: Creativity through emergent binding in neural networks. *Cognitive Science*, *35* (1), 1–33.

Thagard, P., Holyoak, K. J., Nelson, G., & Gochfeld, D. (1990). Analog retrieval by constraint satisfaction. *Artificial Intelligence*, *46* (3), 259–310.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, *53* (1), 1–25.

Tulving, E., & Watkins, M. J. (1973). Continuity between recall and recognition. *The American Journal of Psychology*, *86*, 739–748.

Vertolli, M. O., & Davies, J. (2014). Coherence net: A new model of cognitive generation. In A. Rosa, J. J. Merelo & J. Filipe (Eds.), *Proceedings of the 6th international conference on Evolutionary Computation Theory and Applications (ECTA)* (pp. 308–313). Rome, Italy: IJCCI.

Vertolli, M. O., Breault, V., Ouellet, S., Somers, S., Gagné, J., & Davies, J. (2014). Theoretical assessment of the SOILIE model of the human imagination. In P. Bello, M. Guarini, M. McShane & B. Scassellati (Eds.), *36th International Conference of the Cognitive Science Society* (pp. 3043–3048). Quebec City, QC: Cognitive Science Society.

Von Ahn, L., & Dabbish, L. (2004). Labeling images with a computer game. In E. Dykstra-Erickson & M. Tscheligi (Eds.), *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 319–326). New York: ACM.

Von Ahn, L., Liu, R., & Blum, M. (2006). Peekaboom: a game for locating objects in images. In R. Grinter, T. Rodden, P. Aoki, E. Cutrell, R. Jeffries & G. Olson (Ed.), *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 55–64). Montreal, QC: ACM.

Ward, T. B. (1994). Structured imagination: The role of category structure in exemplar generation. *Cognitive Psychology*, *27* (1), 1–40.

Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, *1* (1), 67–82.

Zhang, M., & Zhou, Z. (2013). A review on multi-label learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, *99*, 1.