Contextual Coherence in the Visual Imagination:
An Interdisciplinary Analysis

by

Michael O. Vertolli

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of
the requirements for the degree of

Masters

in

Cognitive Science

Carleton University

Ottawa, Ontario

© 2014

Michael O. Vertolli

Abstract

This thesis describes an analysis of contextual coherence in the visual imagination using three disciplines: cognition, computation, and neuroscience. I examine the topic by augmenting a model of the visual imagination, SOILIE, with an improved version of their top- n model of coherence. I show that the augmented, serial local hill search model, Coherencer, is an improvement over the original model using a new, quantitative evaluation. I then demonstrate that Coherencer is better than a competitive model from the literature on general coherence; it is better than the original top- n model across different compression representations, mainly co-occurrence probabilities and holographic vectors; and it is consistent with contemporary, neuroscientific research on the hippocampus, specifically Scene Construction Theory.

Acknowledgments

This work would not have been possible without various members of the academic community within and outside of Carleton University. First and foremost, I would like to thank my supervisor, Jim Davies, for his tireless support on all matters related to my academic career.

Thanks to my committee member, Tony White, for his helpful suggestions on this manuscript.

Thanks to all the members of the Science of Imagination Lab for their help and patience during my ceaseless ramblings at the lab meetings. Thanks to Matthew Kelly for his contributions to the manuscript. And finally, thank you to Jeremy Burman for setting me down the path that inspired me to stay in academia. Whatever success I have achieved up to this point is largely the consequence of the overwhelming support of this community, my friends, and my family.

Table of Contents

Abstract.....	ii
Acknowledgments.....	iii
Table of Contents.....	iv
List of Tables	vii
List of Figures	viii
List of Equations	ix
1. Introduction	10
2. Previous Research and a New Model	12
2.1 Introduction	12
2.2 The Coherencer System	17
2.3 Task	18
2.4 Method	20
2.5 Results	21
2.6 Discussion	22
2.7 Conclusions	23
3. Coherence Research	24
3.1 Introduction	24
3.2 Thagard's Model	27
3.3 Task	29
3.4 Method	29
3.5 Results	30
3.6 Discussion	30
3.7 Conclusions	31
4. Compression and Decompression	33
4.1 Introduction	33
4.2 Holographic Vectors	35
4.3 Task	37

4.4 Method	38
4.5 Results	38
4.6 Discussion	41
4.7 Conclusions	42
5. Scene Construction Theory and the Hippocampus.....	44
5.1 Introduction	44
5.2 Parallels with SCT.....	46
5.2.1 An alternative to memory.	46
5.2.2 Modularity.	49
5.3 Differences between Coherencer and SCT	50
5.3.1 Coherence or Spatial Navigation.....	50
5.4 Coherencer's Contributions to SCT	51
5.4.1 Functional loops.	51
5.4.2 A serial process.....	52
5.4.3 The coherence of spatial integration.....	53
5.5 Discussion	54
5.5.1 Other theories.	54
5.5.2 Summary.....	55
5.6 Conclusions	55
7. General Discussion and Conclusions.....	58
7.1 Introduction	58
7.2 Problems.....	59
7.2.1 Parameter search and validation.	59
7.2.2 Overfitting and prediction.	61
7.3 Contributions.....	64
7.3.1 Visual imagination.....	64
7.3.2 Contextual coherence.	65
7.3.3 Decompression.	67
7.3.4 Generative cognition.....	68
7.3.5 Holographic memory.....	68
7.3.6 SCT and the hippocampus.....	69

7.4 Miscellaneous Assessments	70
7.5 Future Research	70
Appendices.....	72
Appendix A Properties of the Peekaboom Database	72
Appendix B Formal Definition	76
References.....	81

List of Tables

Table 1: Label co-occurrence probability of two images alone and in SOILIE's complete database.....	15
Table 2: Co-occurrence matrix of 'bow,' showing the co-occurrence of the label in the column with the label in the row.....	18
Table 3: McNemar χ^2 calculation between Coherencer and top-n.....	22
Table 4: McNemar χ^2 calculation between Coherencer and Thagard's model (TM).	30
Table 5: Logistic regression predictors with significance and odds ratios.	40

List of Figures

Figure 1: Incoherent image that blends two contexts from query ‘mouse’ with output ‘rat,’ ‘ear,’ ‘computer,’ ‘keyboard.’	16
Figure 2: Control in Coherencer	17
Figure 3: Control in Thagard's model	27
Figure 4: Agent success rate with different compression-decompression pairs out of a total 8,372 possible successes.	39
Figure 5: Threshold search for Coherencer on co-occurrence representation	60
Figure 6: Threshold search for Coherence on holographic representation	60
Figure 7: Thagard's model threshold search	61
Figure 8: Histogram of the quantity of labels with a given number of images containing it.	73
Figure 9: Histogram of the quantity of labels with one image with a given number of co- occurring labels.	73
Figure 10: Histogram of the number of labels with a given number of co-occurring labels.	74
Figure 11: Histogram of the normalized betweenness centrality scores for each label.	75
Figure 12: Formal pseudocode for Coherencer.	79

List of Equations

Co-occurrence Probability Equation.....	13
Model Run Metric Equation	20
Node Activation Update Equation	28
Net Activation Change Equation	28
Variable Element-wise Multiplier.....	29
Change in Activation Equation.....	29

1. Introduction

The imagination is implicated in a wide range of abilities related to human cognition. The list of abilities, includes but is not limited to, planning, problem solving, hypothetical thinking, counterfactual thinking, theory of mind, and mental time travel (Davies, Atance, & Martin Ordas, 2011). Despite a plethora of research on the imagination as a facilitator for these abilities (see, for example, Markman, Klein & Suhr, 2012), the generative capacity of the imagination is an untapped area of research. This work focuses on the most studied faculty: vision.

When someone constructs a visual scene with the imagination (e.g., a mouse eating a piece of cheese) they might use visual memories from many different experiences as the components of the new scene. How these components are selected from the range of possible experiences is not obvious. If more than just 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 or appropriate than others (e.g., a rollercoaster, map of Spain, 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 (see Cockbain, Vertolli, Davies, 2014). There is an intuitive coherence imposed on imagined scenes that inhibits unusual and sometimes even highly creative combinations.

One way that humans might make this selection is through the co-occurrence of objects in visual memory (by visual memory I mean only the memory of visual things, and not a specific subsystem like the visuo-spatial sketchpad). Thus, when one is imagining a scene given an environmental query (e.g., a novel, question, or problem), mental processes might search visual memory for other objects that often occur with that query.

When one imagines a mouse eating a piece of cheese, it is not surprising that mousetraps, cats, etc. are more likely to come to mind than unrelated elements. They all often occur together in the world and, thus, are associated in the brain. In this case, “a mouse eating a piece of cheese” serves as the query and the other elements are what are returned by some imagination process. One way to further explore this idea is through the use of computational models.

The paper proceeds over the course of four sections. Each of the four sections expands the topic and report an experimental assessment of the hypotheses of that expansion. The experiments in each section are also based in part on work published in conference proceedings. The papers are indicated where appropriate. Modifications and elaborations on the original papers are included where necessary. The closing general discussion operates as a summative review of all of the work on the topic, thus far.

The next section (2.) addresses previous research as well as introduce and test a new model of contextual coherence: Coherencer. The third section introduces a broader literature on coherence in cognition, specifically the work of Paul Thagard, and compares the corresponding model to Coherencer. The fourth section broadens the scope of the problem by describing Coherencer and the related models as decompression techniques and then testing whether Coherencer is robust across different compression representations, specifically holographic vectors and co-occurrence probabilities. The fifth and final section gives a more detailed description of the associations between Coherencer and Scene Construction Theory: a new theory of hippocampus function. Though this section proposes an experimental assessment, it is left for future research. Contributions are made to the empirical and computational understanding of the visual imagination, contextual coherence, decompression, generative cognition, holographic memory, and the hippocampus.

2. Previous Research and a New Model

This section discusses previous research and motivates the creation of a new model, Coherencer, for resolving contextual coherence in the visual imagination. The experimental results are new, but based on the work of Vertolli and Davies (2013). In this research I designed and implemented the model, and did the write up. Professor Davies acted as primary editor.

2.1 Introduction

The Science of Imagination Laboratory Imagination Engine (SOILIE) is a computational model of a functional description of the generative processes of the imagination (Breault, Ouellet, Somers, & Davies, 2013). In place of human ‘experiences’ and ‘objects’ SOILIE uses labeled images from the web. When generating a novel scene, SOILIE must determine which labels are appropriate to select, given a particular query. That is, given something to imagine, such as “plate,” SOILIE’s task is to choose what other elements should appear in the scene with it (such as “fork” and “dinner”). And, in keeping with the descriptions given above, SOILIE currently uses co-occurrence relations to make this selection. In this context, co-occurrence is determined by the frequency with which one label is present in the same image with another label.

SOILIE derives these co-occurrence relations from the Peekaboom database of labelled images. With over fifty thousand images and ten thousand labels, the Peekaboom database is one of the largest of its kind. The dataset is the combined result of two online games: the ESP Game and Peekaboom (Von Ahn, Liu, & Blum, 2006). In the ESP Game, pairs of players are shown the same image and without communicating try to enter the same words (Von Ahn & Dabbish, 2004). Words that both players enter are associated with the image and, consequentially, common labels are applied to images collected from the internet. To prevent a narrow set of the

most common words, labels would become unusable after repeated use. This increased the diversity and size of the resulting label set for each image.

SOILIE’s dataset comes from a related game, Peekaboom, which uses ESP game data and results in images with labelled selections of pixels. Both games are designed to produce data that can be used in vision research. Thus, they are particularly relevant for SOILIE’s task. For a detailed description of some properties of the Peekaboom database, see Appendix B.

SOILIE uses co-occurrence probabilities as an approximation of conditional probabilities that are derived 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:

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

where $\cap$ 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 value for mouse-cheese than it does for cheese-mouse). Parallel research on co-occurrence in the machine learning literature suggests that this is both more realistic (e.g., almost all weddings have flowers but most flowers are not in weddings) and most models do not account for it (see Huang, Yu & Zhou, 2012; Zhang & Zhou, 2013).

Research in neuroscience suggests that visual working memory can hold approximately three to five objects of “average complexity” (Cowan, 2001; Edin, *et al.*, 2009). Thus, it is assumed that on average an imagined scene has approximately three to five elements in it at any given time; though aggregates (i.e., combining two or more elements into a single element) are entirely possible, I have chosen to ignore them for simplicity. Thus, four labels, excluding the

query, are retrieved by SOILIE from the co-occurrence data and five labels in total are selected for every imagined scene. I decided that this number, despite being in the upper part of the range, was the most useful: preliminary research suggested that larger sets of labels increased the divergence in the success of the underlying subsystems, five is still in the accepted range, and the query does not really need to be maintained in working memory to the same degree (an individual could always re-query) nor does it need to be retrieved.

However, after working with earlier instantiations of SOILIE (see Breault, Ouellet, Somers, & Davies, 2013), a problem became apparent. When images are selected purely on the basis of their co-occurrence with the initial query, that is, selecting the labels with the highest co-occurrence or “top- n ,” 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; e.g., a computer mouse and the animal mouse). Each meaning of a polysemous word is represented by different images in the database—assuming that a single image would be highly unlikely to contain both meanings of a given label (e.g., have the animal on a desk with a computer). Each of these different images is similarly associated with a different collection of labels and each set of labels has a different set of co-occurrence relations. Problematically, by being reduced to a collection of co-occurrence probabilities in visual memory, the sets of images, labels, and co-occurrence relations that separate the two polysemous meanings of the word “mouse” are no longer directly detectable. They are collapsed into a single dimension associating pairs of labels (see Table 1).

Table 1: Label co-occurrence probability of two images alone and in SOILIE's complete database.

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 from the two images using co-occurrence from database:

rat, ear, computer, animal

This problem is not limited to word overlap. Given a particular selection of labels, all of which co-occur, it is still possible that no image exists in the database with that combination. For example, one image might contain “mouse,” “rodent,” “rat,” another might contain “mouse,” “animal,” “ear,” and a third might contain “rodent,” “rat,” “animal,” “ear.” Thus, all five labels would co-occur in the database in general without those labels all being present in any individual image.

Figure 1: Incoherent image that blends two contexts from query ‘mouse’ with output ‘rat,’ ‘ear,’ ‘computer,’ ‘keyboard.’

The result is that models, like SOILIE’s original “top- n ” model, will be unable to infer the appropriate contextual relations from the differences in the underlying images. They take the co-occurrence probability representation as it stands by focusing on only a single dimension: co-occurrence with the query. Thus, they will often produce incoherent images (see Figure 1). In order to ameliorate this problem, I chose to augment the top- n approach with a paired association search using a serial, local-hill searching algorithm. The resulting model is Coherencer.

2.2 The Coherencer System

Figure 2: Control in Coherencer

The new, augmented approach, from now on described as Coherencer, operates as follows (see Figure 2 or Appendix A for a formal description). First, a top-*n* search gathers the initial pool of four labels that co-occur with just the query. Then, an associative search checks the degree to which each label in the pool co-occurs with all the others, as well as with the query.¹ The network of co-occurrence relations that results is tested against a selection threshold. Labels with low co-occurrence in the network are swapped out and new labels that co-occur with the query

¹ The co-occurrence with the query still contributes to the co-occurrence overall, which is why it is included in the search despite previous consideration in the top-*n* model.

are randomly swapped in until the threshold for the network as a whole is exceeded. Once the threshold is exceeded, the set that remains is returned for inclusion in the imagined scene. If the set of labels that co-occur with the query is exhausted, the model returns the best collection of labels with the highest mean co-occurrence probability.

The first part of Coherencer performs a top- n search by selecting the four labels with the highest co-occurrence probability with the query. The second part takes all five labels, including the query, and produces a co-occurrence matrix. Using “mouse” as an example, one would get Table 2.² Each cell in the matrix holds the co-occurrence probability, with the row as the query (q) and the column as the co-occurring label (l) in the co-occurrence probability formula.

Table 2: Co-occurrence matrix of ‘bow,’ showing the co-occurrence of the label in the column with the label in the row.

Labels	Mouse	Rat	Ear	Computer	Animal
Mouse	-	0.29	0.19	0.17	0.13
Rat	0.61	-	0.34	0.00	0.37
Ear	0.04	0.03	-	0.01	0.12
Computer	0.04	0.00	0.02	-	0.00*
Animal	0.03	0.04	0.14	0.00*	-

*probabilities were greater than 0 but required too many digits

2.3 Task

The problem of ‘coherence’ is not exclusive to SOILIE.³ Models that address context need to be able to select coherent combinations (Hullett & Mateas, 2009). Most models, due to memory

² Though the co-occurrence probability value of a label with itself should actually be 1 mathematically, it was convenient to ignore the values in the context of the implementation.

³ For a formal definition of the problem, see Appendix A.

limitations, will also need to reduce the original input (e.g., images) into some form of compressed data (e.g., co-occurrence probabilities). In this compression, some of the information will be lost. Thus, one can see the problem, contextual coherence in the visual imagination, as a member of the general class of problems dealing with lossy compression. This topic will be addressed in detail in Section 4; however, it is this basic idea that motivated the current task design.

The comparison followed the basic structure of a memory task, where a participant is given a collection of data; this data is compressed in memory, and then it is recalled. In the abstraction of this structure, each model is given a collection of images, which are compressed into co-occurrence probabilities in memory. The models are then tasked with recalling this information.

Unlike a memory task, the goal is not to test the bounds of human or animal functionality. Instead, it is to assess the efficacy of the decompression step that deduces the coherence information that was lost during compression. Thus, the quantity of images remembered is not what is of interest. It is a certain quality in the generated images, mainly coherence. This quality can be tested quantitatively by determining if the elements (in this case, labels) selected by the model when given a particular label or query do in fact occur in one of the original images. If they do, then the original coherence information has been successfully deduced from the compressed data.

To test Coherencer's efficacy, I compared it to SOILIE's original top- n model. The hypothesis was that Coherencer will outperform the top- n model.

2.4 Method

There are two models that were compared: Coherencer and the top- n model. The top- n model retrieves the four labels with the highest co-occurrence probability with the query.

The entire Peekaboom database was initially filtered to remove all images with fewer than five labels and any labels that only occurred in those images. A total of 8,372 labels and 23,115 images remained after this filtration. All of the remaining images were compressed to their corresponding co-occurrence probabilities.

Each of the 8,372 labels was run through both of the models. The top- n procedure always yields the same result (the top-4 associated labels); thus, the model was run once per query label. Coherencer has stochastic variation in its results; thus, it was run 10 times on the entire set of 8,372 query labels. The total results for the entire set were averaged across all runs.

The number of runs conforms to Byrne's (2013) analytic model run metric that specifies the necessary number of runs to achieve a statistically robust computational result. This metric is derived from the formula for confidence intervals for proportions. This metric uses the following formula:

$$n = p(1 - p)(z/w)^2 \quad 2$$

where n is the necessary number of model runs, p is the given proportion for the model, z is the value of the normal distribution for the upper tail of the desired confidence interval distribution, and w is half the proportional 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.

Each query plus four returned labels are the elements of a new, generated image. The results for each of the models were assessed with regard to the original images. If at least one of

these images contained the five labels that were selected by a particular model, including the query, the model scored one point. If there were no images containing the five labels, it did not score a point. The total number of points scored by a model where the other model failed to score a point (i.e., excluding labels where both models failed or both models succeeded) were used for comparison.

2.5 Results

As hypothesized, Coherencer had more successful matches than the top- n model.

McNemar's repeated measures chi-square test demonstrates that Coherencer performed significantly better than top- n , $\chi^2(1, N=8372) = 2602.66, p < .000, \phi = .22$. The average scores in each of the categories are listed in Table 3. In this test, model runs where Coherencer and top- n both fail or both succeed on a given query (i.e., the models perform identically) are ignored; thus, the comparison occurs between the runs where one model failed and the other succeeded and vice versa. All values are reported for completion and evaluation purposes. 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.

Table 3: McNemar χ^2 calculation between Coherencer and top- n .

		Coherencer failure	Coherencer success	Total
Top- n failure	Actual	535.0	2995.0	3530.0
	Expected	284.2	3245.8	
Top- n success	Actual	139.0	4703.0	4842.0
	Expected	389.8	4452.2	
Total		674.0	7698.0	8372.0

2.6 Discussion

The results support the idea that Coherencer generates scenes that are more coherent than the top- n model. This adds further support to the trend found in Vertolli and Davies (2013), which also found that top- n was better than a random selection control. For Coherencer, this fact is all the more impressive, given how many challenges were stacked against it. Partially, these challenges were due to lexical confounds. For example, synonyms confine the search space (e.g., when one searches ‘dog’ it will not include the related associations for ‘puppy’), confound the output (e.g., by returning ‘puppy’ when searching for ‘dog’), or result in false negatives (e.g., ‘puppy’ is included in an image with the other labels but ‘dog’ is not; the match fails as a result).⁴ Hyponyms and hypernyms (e.g., ‘dog’ to ‘German Shepherd’ and ‘German Shepherd’ to ‘dog,’ respectively) result in similar problems as do meronyms (e.g., ‘nose’ to ‘face’) and other lexical relations. These problems were left for future research.

⁴ I would argue that these challenges do not completely generalize to the top- n model. For example, increasing the search space would increase the interference or noise for the top- n model while it increases the possibility of finding a better match for Coherencer. However, improvements that restrict the search space *might* generalize.

2.7 Conclusions

A number of contributions have been made in this section to the study of contextual coherence in the visual imagination. First, I have provided an initial description of the problem: compression representations (e.g., co-occurrence probabilities) that do not explicitly represent context relations will need more complex decompression techniques like Coherencer. Second, there is a new (at least in the imagination and coherence literatures), quantitative method for assessing contextual coherence: awarding points for selected label sets that occur in an image in the uncompressed database. The key here is that, if context is lost in the process of compression, its deduction is a measurable achievement. Finally, there is preliminary empirical support for the notion that Coherencer is a better system for deducing contextual coherence in the visual imagination than the top- n model on the Peekabom dataset. Consequently, I can conjecture that future computational models of the visual imagination could use Coherencer as an empirical and computational benchmark in this domain. However, it is unclear how good Coherencer is at generating coherence in domains other than the visual imagination or in terms of more robust models. This led to the work of Paul Thagard (2000).

3. Coherence Research

This section discusses Paul Thagard's (2000) description and computational model of coherence. It then compares his model to Coherencer. The experimental results are new, but based on the work of Vertolli and Davies (2014). In this research I implemented Thagard's model, and did the write up. Professor Davies acted as primary editor.

3.1 Introduction

Paul Thagard (2000) devotes an entire book to the coherence problem, broadly construed. In this text he describes coherence as an optimization problem. That is, coherence is the dynamic selection of the best combination of elements to maximize or minimize a particular 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. Note that, for Coherencer, negative constraints are implied by low co-occurrence probabilities between pairs of labels.

After formalizing coherence in this way, Thagard proceeds to outline a number of general classes of computational models that can resolve this type of problem. One of the first algorithms that he describes is the class of "incremental" algorithms. Like Coherencer, incremental algorithms evaluate the coherence of a single element at a time relative to the current pool of selected elements. However, there are a number of differences.

First, the incremental algorithm builds its initial pool one element at a time whereas Coherencer seeds its initial pool with the top- n model. Second, the space within the incremental algorithm's "working memory" can be of arbitrary size: it could literally contain the entire set of

possible elements if that was what maximized coherence. In contrast, Coherencer has a finite limit on the size of its pool, limited by constraints on human working memory. Third, Coherencer does not *maximize* coherence; it only makes sure that it passes a certain threshold (λ). Fourth, once an element has been selected by the incremental algorithm, it cannot be unselected (i.e., there is no backtracking with selected elements). Coherencer maintains backtracking capabilities for selected elements (i.e., it can discard what it has already selected); however, both Coherencer and the incremental algorithm cannot backtrack when they reject an element as incoherent. Despite these differences, I take the Coherencer model to be under the same class and ‘incremental’ works as a label for this class: it highlights the serial approach that is a defining feature.

After defining the original, incremental algorithm, Thagard proceeds to argue that this class is problematic, at least prescriptively (i.e., for indicating the ideal). They often lead to suboptimal solutions (i.e., they are liable to get stuck on local optima). Largely, this is a result of the serial increments by which they make comparisons, a property that I take to be central to the class. That is, since Coherencer and similar models can only examine the relations of the current element, rather than all possible elements in parallel, they are more likely to get trapped on local optima.

Coherencer, for example, might remove ‘computer’ (lowest total row and column co-occurrence of 0.24; see Table 2 in Section 2), even though ‘computer’ might actually be a part of the set with the highest possible average co-occurrence for the query. Rejecting ‘computer,’ which is reasonable at this stage, prevents the model from re-attaining it later or even finding the ‘best set’ of which computer is a part. Though backtracking for these rejected elements has been

implemented as a fix for serial approaches in general, Thagard suggests that this is still worse than some alternatives.

In response to this problem, Thagard (2000) proposes a connectionist model that is not hindered by serial processes. These models examine possible solutions in parallel, which decreases the possibility of getting stuck on sub-optimal solutions. As a result, it *should* outperform Coherencer. Thagard has implemented a number of connectionist models with success (e.g., Thagard, 1989, 1991, 1992a, 1992b, 2000; Eliasmith & Thagard, 1997; Thagard, Holyoak, Nelson & Gochfeld, 1990). However, before continuing his exploration of his preferred approach, Thagard makes one caveat; mainly, that incremental algorithms, or what are commonly referred to as local hill searchers in the machine learning and optimization literature, might offer valuable insights into human cognition: both are known to perform sub-optimally in many domains, including coherence.

In what follows, SOILIE's Coherencer system will be compared to one instantiation of Thagard's connectionist algorithm in the current domain (i.e., visual coherence in the human imagination). In Section 2, Coherencer was initially shown to better capture coherence than SOILIE's original, top-*n* model in an experimental evaluation. It was also shown that Coherencer falls under Thagard's incremental class of algorithms. Thus, this comparison provides both a more robust test of Coherencer's efficacy, and insight into the cognitive implications that Thagard pointed to in his caveat.

It is worth noting that both Coherencer and Thagard's model, in as much as they exist in the brain, are *both* necessarily instantiated in neural processes. One should not confuse the semantic convenience of calling connectionist models "neural networks" with a literal neural network and Coherencer with "something else." What is being tested, then, is whether the

higher-order functionality of neural processes is better replicated with a *serial* process or a *parallel* process, with the corresponding implications for optimality in the system. A serial virtual machine can be implemented on a parallel computational architecture, neural or otherwise. Thus, both types of processes are cognitively plausible when considering only this aspect.

3.2 Thagard's Model

Figure 3: Control in Thagard's model

The construction of the connectionist model will proceed as described by Thagard (2000) (see Figure 3). A node is constructed for the query and every label co-occurring with the query. For

every positive constraint between two labels, an excitatory link is constructed between the corresponding nodes with a weight equal to the co-occurrence probability. For every negative constraint, an inhibitory connection is constructed between corresponding nodes with a weight set to the average co-occurrence of all non-zero values (π).⁵ An initial activation (0.01) is assigned to each node with a special locked activation (1.0) on the query node. All nodes then have their activation updated in parallel⁶ using the following formula⁷:

$$\vec{a}_{t+1} = \vec{a}_t(1 - d) + f(\overrightarrow{net}) \quad 3$$

where $\vec{a}$ is a vector of all the node activations at time t , d is a scalar decay parameter (0.05) that decrements each node at every cycle. The vector $\overrightarrow{net}$ is computed by:

$$\overrightarrow{net} = \vec{a}_t W \quad 4$$

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., neighbours 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 π . Finally, f from the original equation is a function that performs element-wise multiplication with a different number depending on the elements direction from zero as per this equation:

⁵ This was found to be 0.14878295850321488. The rounding occurred where it naturally does in the Python computer language (double precision float). As a consequence, different languages may get slightly different results unless this is controlled.

⁶ The nodes are updated serially but the results of those updates are not used until the next serial update of all nodes. Thus, the end product is a parallel process implemented on a serial machine.

⁷ All formulas are vectorized implementations of those described by Thagard (2000). I chose to use row vectors instead of column vectors as this more closely mirrors Coherencer's implementation.

$$f(\overrightarrow{net}) = net_i x \begin{cases} x = a_{max} - a_i & \text{if } net_i > 0 \\ x = a_i - a_{min} & \text{if } net_i \leq 0 \end{cases} \quad 5$$

where x is the variable multiplier, a_i is the i th value of $\vec{a}$, 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 reduced 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 (θ) or until 500 iterations occur. The following equation illustrates the former:

$$\Delta \vec{a}_t = \frac{1}{10n} \sum_{t=10}^t \sum_{i=0}^n (|a_{t,i} - a_{t-1,i}|) < \theta \quad 6$$

where Δa 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 θ is the threshold of 0.005. The 4 labels with the highest activation are selected providing a sort of top-4 filter.

3.3 Task

The current task follows the same outline as the task in Section 2.2. The hypothesis is that Coherencer will outperform Thagard’s model in the current comparison. I anticipate that serial processes better capture the contextual transitions necessary to appropriately frame a given scene. And, the advantages of using a parallel, non-linear optimization process are lost when dealing with a single feature.

3.4 Method

The current method follows the same outline as the method in Section 2.3. However, due to a greater similarity in the success rates of Coherencer and Thagard’s model, Byrne’s (2013) metric required 39 model runs for each model.

3.5 Results

As hypothesized, Coherencer had more successful matches than the top- n model.

McNemar’s repeated measures chi-square test demonstrates that Coherencer performed significantly better than top- n , $\chi^2(1, N=8372) = 2308.75, p < .000, \phi = .23$. The average scores in each of the categories are listed in Table 4. In this test, model runs where Coherencer and Thagard’s model both fail or both succeed on a given query (i.e., the models performed identically) are ignored; thus, the comparison occurs between the runs where one model failed and the other succeeded and vice versa. All values are reported for completion and evaluation purposes. 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.

Table 4: McNemar χ^2 calculation between Coherencer and Thagard’s model (TM).

		Coherencer failure	Coherencer success	Total
TM failure	Actual	510.0	2751.0	3261.0
	Expected	260.6	3000.4	
TM success	Actual	159.0	4952.0	5111.0
	Expected	408.4	4702.6	
Total		669.0	7703.0	8372.0

3.6 Discussion

The results support the idea that Coherencer generates scenes that are more coherent than Thagard’s model on the given dataset. This offers further support to the results found in Vertolli and Davies (2014). However, the intent in both cases is not to falsify Thagard’s claim to the formal optimality of connectionist algorithms over incremental algorithms in the domains he

considers: largely higher-order epistemological relations and constraints. Co-occurrence probabilities are part of a much lower system. The fact that Thagard's theory could anticipate both categories of systems I believe lends credence to it.

The purpose of this comparison is to extend the theory in order to better comprehend the subtle nuances implicated within it. For example, under what conditions are incremental algorithms cognitively useful? Here, the evidence suggests that incremental, heuristic approaches function well with simple representations (i.e., just co-occurrence), low dimensionality (i.e., just co-occurrence probabilities), with high combinatoric load (approximately 3.42×10^{17} possible 5-label combinations given the 8372 possible labels). Assuming the connectionist, parallel approach is optimal as Thagard (2000) described, how might that incremental approach switch into a functionally parallel one? Or, does it only approximate a parallel approach, which forces sub-optimal solutions in higher-order domains? Thagard (2000) explicitly mentions the tendency for humans to make sub-optimal decisions as well as the potential association between incremental approaches and bounded rationality (Simon, 1991). This project supports this association.

3.7 Conclusions

A number of contributions have been made in this section to the study of contextual coherence in general and in the visual imagination. To begin, I can expand my conclusions from the previous section (2.7). Thus, I have support for the notion that optimization techniques as defined by Thagard, specifically incremental optimizers like Coherencer, perform well as decompression procedures for lossy forms of compression. The reverse claim that coherence *in general* can be viewed as a compression-decompression problem is a suggestive conjecture but requires further evidence. I will withhold its discussion in detail until the next section. Additionally, there is

further support that Coherencer is a useful empirical and computational benchmark for research on the visual imagination *and*, potentially, coherence broadly. Finally, the quantitative evaluation task does not generalize as well to coherence problems, broadly, unless they can be mapped to the compression-decompression framework; mainly, they need to have known real world examples of coherence that can be tested against and the property of optimality or coherence must be lost in the stored representation.

On the basis of the current comparison, there is support for the notion that parallel optimizers (e.g., Thagard's model) are not guaranteed to be more optimal than incremental optimizers (e.g., Coherencer) purely on the basis of their parallel-incremental distinction. Second, I can also tentatively conjecture that domains with simple representations, low dimensionality, and high combinatoric load might be better suited to incremental algorithms. Future research is necessary to lend further support for this claim, but these properties were consistent with the Peekabom dataset.

Given the importance of the compression-decompression framework for the current research, I decided it would be a good place for further research. The next section discusses this research in detail.

4. Compression and Decompression

This section discusses my mapping of the contextual coherence problem onto a compression-decompression framework in detail. It then reports an expansion of the original test from section 2 comparing the top- n model and Coherencer to a new compression technique: holographic vectors. The experimental results and exposition are based on the work of Vertolli, Kelly, and Davies (2014). In this research I implemented the holographic vector representation, and did the write up on everything except the holographic vector section. Professor Davies acted as editor. Kelly wrote the holographic vector section and acted as editor.

4.1 Introduction

Compression has been implicated in artificial general intelligence and, more broadly, general cognition (see Hutter, 2005; Schmidhuber, 2009; Wolff, 2013). In the context of vision, for example, there is massive redundancy in the information presented by natural images (Attneave, 1954; Ruderman & Bialek, 1994). Thus, one of the main functions of the early visual system is to reduce this redundancy (Attneave, 1954; Barlow, 1961). These reduced representations are then transformed into invariant representations of objects later in the processing stream (Rolls, 2008). Though research in this literature has not always focused on compression, it is undeniable that the compression of information from the environment is critical to an organism's success (Simoncelli & Olshausen, 2001). Thus, compression, understood as the reduction and abstraction of irrelevant details of a stimulus to those that are relevant for the survival of the organism, is an important mechanism in cognition.

Given this framing, one can see the successful functioning of a given agent or organism on a continuum of optimality, with perfect compression of all possible stimuli on the successful

end and compression of no stimuli on the failing end.⁸ Every agent will fall somewhere on this continuum. What this means is that, barring the optimal limit, every agent is engaged in a lossy form of compression to some degree. Presumably, species evolve and individual agents develop to have compression mechanisms appropriate for their flourishing. This can be viewed as an optimization procedure.

I propose this optimization procedure as the interaction of two related but differentiable tasks. First and foremost, there is the primary task of achieving better compression techniques that capture more of the information in the environment. Second, there is the task of achieving better decompression procedures that can re-derive information that was lost through higher-order patterns in the compressed encodings. The first task (compression) is captured by the full range of machine learning techniques that move from instances of a phenomenon (e.g., pictures) to general classes or regularities, much like inductive inference. The second task does the reverse: it moves from general, often overlapping classes or regularities to instances, much like deductive inference. In framing the problem in this way, part of the contribution is to demonstrate that decompression, which has received much less attention, is a task amenable to techniques commonly used in artificial intelligence research.

The reason I propose this division of the optimization of the agent is because the compression representation *on its own* can never fully define the actions and interactions of the agent. This is a reduction of the complexity of the agent to its structure to the exclusion of the interactions of the agent *with* that structure. I argue that this reduction, though useful in the early stages of a research program, is a false reduction: it excludes relevant content for the accurate

⁸ One will note that there is another, related dimension to optimality that has to do with the amount of noise produced in the compression or the degree to which information is produced that is not representative of the input. The intricacies of the relationship between these two dimensions are complex and largely outside the scope and interest of this work. Thus, I have chosen to ignore the noise dimension in this section.

and complete description of the agent. In the visual imagination, this fact is self-evident: *on their own*, co-occurrence probabilities are meaningless. A minimal decompression technique like the top- n model is necessary to use the given structure.

In the context of cognition generally, I have chosen to describe decompression as a form of generative cognition. By generative cognition, I mean the production of new properties and relations from a given data set that are not explicitly stored in the data set. When an agent imagines a new scene on the basis of an environmental trigger with elements that were not explicitly encoded in the memory or in the trigger (e.g., contextual coherence), the agent is engaging in generative cognition. Generative cognition is distinct from creativity in that the product of a generative, cognitive process can be mundane or uncreative.

In what follows, I will evaluate the original comparison of the top- n model and Coherencer from Section 2 using two compression techniques: co-occurrence probabilities and holographic vectors.

4.2 Holographic Vectors⁹

The new compression representation I chose to compare with co-occurrence probabilities is the holographic vector or holographic memory representation. The justification for this comparison is that holographic vectors have been used to specifically handle context-based information in text (Jones & Mewhort, 2007). Thus, though tested in a slightly different domain, it seems to be a representation that is likely to capture more salient context information than simple co-occurrence probabilities. At minimum, the structure of the holographic vector representation should capture *different* information.

⁹ This section was written almost entirely by Matthew Kelly in the associated paper (Vertolli, Kelly, & Davies, 2014). It has been included for completeness with his permission with a few modifications.

The holographic memory representation is setup as follows. Each label in the Peekaboom database is represented by a vector of 1000 dimensions. Each of these vectors is generated randomly by sampling 1000 values from a normal distribution with a mean of zero and a standard deviation of $1/\sqrt{1000}$. As a result, each vector has a Euclidean length of approximately one and is approximately orthogonal to every other vector. These vectors are termed by Jones and Mewhort (2007) *environment vectors*.

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 I adopt the simplest technique used by Jones and Mewhort (2007). For the purposes, an image is a collection of co-occurring labels. The memory vector for a given label is the sum 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's (A) 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.

Similarity in a vector space model is measured as the angle between vectors. Typically, the vector cosine is used. The cosine ranges from 1 to -1. When comparing a pair of vectors, a cosine of 1 indicates that the pair of vectors has an angle of zero degrees and they are identical representations. A cosine of -1 indicates that the pair of vectors has an angle of 180° and they represent exact opposites. A cosine of 0 indicates that the pair is at a 90° angle and they are 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 $P(l \mid q)$, and as such will be less accurate than storing the exact co-occurrence probabilities, so there is no particular advantage to using vectors in this way when one can instead use exact co-occurrence probabilities.

The advantage of using holographic vectors comes from (1) their ability to represent arbitrarily complex associations in a compressed form using convolution and (2) their spatial nature, allowing for easy comparisons between any two representations to be made. With regards to (1), representing associations between groups of labels is beyond the scope of this paper.

One can take advantage of (2), the spatial nature of the vectors. The cosine of a memory vector of one label and an environment vector of another label is a measure of first-order association: how often the labels appear in *the same* image. The cosine of a memory vector with another memory vector is a measure of second-order association: how often the labels appear in *similar* images. When processing language, first-order association indicates how likely two words are to appear in the same sentence, whereas second-order association is indicative of synonymy. Using second-order association to construct coherent scenes is an interesting alternative to using co-occurrence probabilities (first-order association). I explore this alternative in what follows and compare task performance for holographic vectors using second-order association to task performance using exact co-occurrence probabilities.

4.3 Task

The current task follows the same outline as the task in Section 2.2. The hypothesis is that Coherencer will outperform the top- n model across both compression techniques, following the

results of Section 2.5, and the holographic vectors, by capturing more contextual data, will outperform the co-occurrence probability representation across both decompression techniques.

4.4 Method

The current method follows the same outline as the method in Section 2.3. However, due to the complexity of the comparison across 4 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. Greater difference between the conditions meant that Byrne's (2013) metric only required 10 model runs for Coherencer on the conditions using co-occurrence probabilities and 9 runs on the holographic vector conditions. Top- n was run once for each compression representation.

4.5 Results

As hypothesized, Coherencer outperformed the top- n model across both compression techniques. Contrary to the original hypothesis, the co-occurrence probability representation outperformed the holographic vector representation across both decompression procedures. The success rates out of the 8,372 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 Figure 4.

Figure 4: Agent success rate with different compression-decompression pairs out of a total 8,372 possible successes.

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), I needed to use a logistic regression analysis in order to assess statistical significance. As this method is less common in the literature, I start by giving a brief description of it.

In a logistic regression analysis, a statistical model of the overall success rates across all conditions is created using the two independent variables, in this case compression and decompression, and their interaction as predictors. The test then assesses whether the model's ability to accurately classify the data is significantly better than a constant only model (e.g., predicting only failures). The effect of each predictor alone is then compared using the Wald criterion while the direction and importance of the predictors is described using odds ratio (e^{β}) values.

The model generated by the logistic regression using the three predictors was able to predict success or failure on the basis of those predictors with 69.8% accuracy overall (73.2% for

success and 75.7% for failure). A test of the full model against a constant only model was statistically significant, indicating that the predictors as a set reliably distinguished between success and failure of the model, $\chi^2(3, N=33218) = 14698.72, p < .000$, Nagelkerke's $R^2 = .47$. The Wald criterion demonstrated that the compression and decompression predictors made a significant contribution to the accuracy of the model (see Table 5). The e^{β} values indicate that the conditions using co-occurrence probabilities and those using Coherencer were more likely to achieve success. The effect of the compression representations was approximately twice as important for the overall effect as the effect of the decompression procedures; the effect of the decompression procedures was approximately eight and a half times as important as the interaction; and, the effect of the interaction was approximately twelve times as important as the constant.

Table 5: Logistic regression predictors with significance and odds ratios.

Predictor	β	SE	Wald	df	p	e^{β}
Compression (Co-occurrence=1, Holographic=0)	2.84	.05	3619.46	1	.000*	17.18
Decompression (Coherencer=1, Top- n =0)	2.10	.05	1965.66	1	.000*	8.17
Interaction (Co-occurrence and Coherencer)	.03	.07	.01	1	.699	1.03
Constant	-2.53	.04	3662.63	1	.000*	.08

* p -values were less than the given score.

4.6 Discussion

The results support the notion that better decompression techniques (e.g., Coherencer) improve the optimality of the agent regardless of the compression technique used (e.g., co-occurrence probabilities, holographic vectors). In this case, co-occurrence probabilities were an improvement over the holographic memory representation and Coherencer was an improvement over the top- n control. This provides an experimental approach designed to separate the effects of compression and decompression on the optimality of the agent while accounting for both. However, the overall size of the effect was quite small, so these conclusions are given tentatively.

The findings continue to support the notion that Coherencer is a better decompression procedure than the top- n approach. However, they contradict the expectation that the holographic vector representation would be better able to capture contextual information than co-occurrence probabilities. Since the goal was mainly to support the differentiable structure of compression and decompression techniques, the analysis required to fully understand why the holographic vector representation did poorly will be left to future research.

Though the task of the current study is restricted to vision following research on the imagination, I take this as a preliminary but illustrative example of a more general property of intelligence and, even more broadly, cognition. Compression is ubiquitous in cognition and the relative optimality of that compression is essential to the success of the agent. This optimality is not given by the mechanisms that perform the necessary reductions in information without the additional support of the mechanisms that can deduce relations that were lost in that reduction. This research is a preliminary demonstration that the exclusion of either side of this dyad necessarily gives an incomplete description of the general process. Thus, I predict that in all

cognitive domains requiring reductive mechanisms, research on better decompression techniques should yield an overall improvement in the system.

4.7 Conclusions

A number of contributions have been made in this section to the study of contextual coherence in general and in the visual imagination. To begin, I can expand the conclusions from the previous section (3.7). First, there is further support for the notion that optimization techniques (e.g., Coherencer) perform well as decompression procedures for lossy forms of compression (e.g., co-occurrence probabilities) relative to the current evaluation scheme. Second, there is further support that Coherencer is a good empirical and computational benchmark for the visual imagination and contextual coherence across *compression* representations. Third, given how poorly the holographic vector representation performed on the current task, I can tentatively conjecture that the quantitative evaluation of contextual coherence is more challenging than the evaluation techniques used in the holographic literature. I can tentatively conclude that this might mean that the variations on the evaluation can be used as a more stringent testing mechanism in related domains (e.g., the holographic vector research on texts).

Some conclusions on decompression, holographic vectors, and generative cognition can be added to these contributions. First, in Section 3.7, I claimed that coherence *in general* could potentially be viewed as a compression-decompression problem. If one accepts the characterization of compression as the reduction and abstraction of irrelevant details of a stimulus to those that are relevant for the survival of the organism and decompression as a mechanism that deduces implicit patterns in the compressed representation, then coherence in general is a compression-decompression problem when it is not explicitly encoded. The generalization of the previous work to domains associated with holographic symbolic

architectures adds further support by showing that coherence is improved and thus not explicitly encoded, or at least well defined, in these two representations. Second, I can conjecture that better decompression techniques will improve the optimality whenever compression is lossy (i.e., important details are not explicitly encoded). Thus, decompression *should* be considered when working in domains using reduction of this kind. Further research is needed to legitimate this claim. Third, the low success rate of the holographic vector representation compared to co-occurrence probabilities across both decompression techniques is both provocative and concerning. At minimum, it suggests that there are some domains where the representation does especially poorly, including the current one. It also could suggest that the current tests in the literature for the representation might be too easy rather than just different. Finally, since decompression is a process that is more general than its current application in the imagination, one can speculate that this mechanism might preliminarily specify generative cognition more broadly. In this way, it might inform other cognitive domains. .

Given the cognitive implications of the research thus far, I decided to focus my research on the underlying neuroscience. In this way I could better integrate the functional-computational descriptions of the visual imagination and generative cognition, broadly, in terms of their biological mappings. This led to the contemporary, neuroscientific research called Scene Construction Theory.

5. Scene Construction Theory and the Hippocampus

This section proposes Coherencer as a functional model of the hippocampus that is consistent with recent research in Scene Construction Theory (SCT). It describes a number of contributions the model lends to the theory and proposes an experimental evaluation of one of them. In this research I did the write up. Professor Davies acted as primary editor.

5.1 Introduction

“Scene construction theory” (SCT) describes a new theory for the role of the hippocampus as a facilitator in the creation of a “coherent spatial context” on which event details of episodic memories and imagined future experiences can be “manipulated, bound and played” (Hassabis & Maguire, 2007; Maguire & Mullally, 2013). In this modality, the hippocampus operates as an anticipatory structure that provides a cohesive spatial framework for what is outside the direct field of experience. One example is when one is aware of the back of the chair that one is sitting on when not touching or visually observing it.

The purpose of this shift is to ground the processes of the hippocampus at a lower, functional level and thereby increase the range of cognitive functions in which it can play a role. For example, the hippocampus has recently been implicated in areas as broad as spatial navigation, future-thinking, and imagination, in addition to its traditional role in episodic memory (Addis, Pan, Vu, Laiser, & Schacter, 2009; Botzung, Denkova, & Manning, 2008; Szpunar, Watson, & McDermott, 2007; Maguire & Mullally, 2013). It is not that the hippocampus is the only area of the brain responsible for these cognitive functions. Rather, according to SCT, it provides an essential ingredient to all of them: the ability to create coherent, spatial scenes.

Though there has been some debate in the literature on this topic (see Hassabis, Kumaran, Vann, & Maguire, 2007; Maguire & Hassabis, 2011; Squire, et al., 2010; Squire, McDuff & Frascino, 2011), the most recent research on boundary extension leans in favour of SCT (Chadwick, Mullally & Maguire, 2012; Mullally & Maguire, 2013). 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) and an object as part of a scene.

Kumaran and Maguire (2009) emphasize a process-based, non-modular computational approach, particularly between memory and perception. However, SCT is a coarse account of the empirical data, and does not explicitly describe the information processing of the underlying neuroscience. An information processing account would allow for the generation of more detailed and specific hypotheses.

The compression-decompression framework described in Section 4 offers an entry point for this information processing account. Thus, one can distinguish learning and generation phases, respectively, as a compression phase, which encodes things into memory, and a decompression phase, which deduces content that was lost in this encoding. In this view, mnemonic recall¹⁰ is just accessing the compressed memory representation and passing it along for future processing. The imagination then takes the compressed outputs of mnemonic recall and deduces what was lost through an optimization of the remaining features in the compressed content that are indicative of higher order patterns.

¹⁰ In order to differentiate this type of ‘recall’ from the standard, psychological sense, I will continue to refer to this type as ‘mnemonic recall.’ If ‘mnemonic’ is absent, assume it is the standard sense.

Though one could argue that the decompression phase, in as much as it contributes to the standard experience of memories, is really just a part of the memory processes (i.e., mnemonic recall), I believe it is not useful to do so. As will be explained in the sections that follow, the generative processes that are implicated in this decompression step are necessary for various cognitive faculties outside of what is standardly viewed as memory (e.g., imagination, future thinking, spatial navigation). Thus, SCT and I view imagination to be on a continuum with normal mnemonic recall.

5.2 Parallels with SCT

5.2.1 An alternative to memory. Hassabis and Maguire (2007) make an effort to differentiate the sub-processes of the network of brain areas associated with episodic memory. This approach occurs on three fronts. First, they firmly situate themselves on the constructive side of the debate on episodic memory (see Conway & Pleydell-Pearce, 2000; Rubin, Schrauf, & Greenberg, 2003; Schacter & Addis, 2007).

5.2.1.1 *Constructive memory hypothesis.* The constructive memory hypothesis states that the recollection of past experiences is a generative reconstruction of experiences on the basis of a trace in long-term memory. This is in direct contrast to theories that propose memories are stored and retrieved in their entirety (for example, see Brewer & Dupree, 1983). The authors argue that the constructive view is more plausible on computational grounds (i.e., due to storage constraints and the needs of generalization and abstraction) as well as cognitive, in that it accounts for well-known memory errors.

Coherencer supports this view by drawing a conceptual delineation between cognitive processes associated with compression and storage (mnemonic), and those that are associated with decompression and deductive reconstruction (generative; Vertolli & Davies, 2014; Vertolli,

Kelly, & Davies, 2014; Vertolli, Breault, Ouellet, Somers, Gagné & Davies, 2014). Coherencer is a model of the generation process; thus, given a particular set of components that were originally compressed in memory (i.e., co-occurrence probabilities; but not the entire scene) and a way to access those components (i.e., mnemonic recall), Coherencer deductively generates the appropriate object relations for the reconstruction of a coherent scene in the imagination. Effectively, Coherencer is one possible computational implementation of the reconstructive functionality described by Hassabis and Maguire.

The component view of reconstruction is also particularly suited for optimization-based formalizations like Thagard's (2000). Or, to describe reconstruction as a combinatorial arrangement of components makes it amenable to processes that function as a form of optimization, even if this optimization is non-optimal (see "satisficing" in Simon, 1956). Thus, not only does Coherencer support the division preliminarily proposed by Hassabis and Maguire (2007), it also offers a more detailed description of the information processing of one side of this division.

5.2.1.2 Neural differentiation. For further evidence for the differentiation of brain areas, Hassabis and Maguire (2007) show a consistent pattern of activation across imagination and scene construction phenomena. In a conjunction analysis comparing 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, Kumaran, & Maguire, 2007). This was found in direct contrast to tasks done in control conditions that required no scene construction (i.e., real and imagined simple objects). 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).

The differentiation of structure-function pairs marks the third part of the argument. Partially, this serves to replace the leading alternatives for hippocampal function. For example, Hassabis and Maguire (2007) show that the subjective sense of time, a sense of selfhood, and auto-noesis (i.e., mental time travel and related phenomena) are all more closely associated with other brain areas despite claims that position them as the primary function of the hippocampus (see Tulving, 2002, for one of these competing views). Partially, the authors use this structure-function separation to further support their claim that the hippocampus is primarily anticipatory as previously described.

In parallel with SCT's view of the hippocampus, Coherencer's primary processes are a form of combinatoric optimization that is directly anticipatory. Or, they answer the question, given x , y , and z , what other objects are likely to be present. The consequence of this is that Coherencer, like SCT's description of the hippocampus, is more general than its current application in the visual imagination and more general than its implementation using co-occurrence probabilities as previously described.

Similarly, though SCT does not explicitly use co-occurrence probabilities in their description of the hippocampus, they do use spatial properties like proximity. Co-occurrence can be thought of as a highly compressed representation of many of these spatial properties as almost every spatial relation *implies* that the objects possessing that relation co-occur. Thus, co-occurrence probabilities as a representation are entirely consistent with the theory.

5.2.1.3 Processing sequence. Coherencer and SCT have a similar description of the processing sequence. For the hippocampus, the construction of anticipated extensions of the

current scene is propagated, in a top-down fashion, back down the network to change the neural firing of the visual cortex towards the extended view rather than the original input (Chadwick, Mullally, & Maguire, 2013). This then propagates back to the hippocampus in a feedback loop.

Coherencer does something similar, if at a much higher functional scope: an anticipated combination of objects (corresponding to a pattern of neuronal firing in the visual cortex) is mnemonically recalled from a bank of possibilities in memory; their co-occurrence probabilities are assessed for coherence (corresponding to hippocampal processing); if it is not coherent enough, one of the selected objects are swapped for a different possible object (i.e., neuronal pattern in the visual cortex).

An example that illustrates the parallel is moving through a dark room. If one knows one started in one's bedroom, it would be highly unlikely that one would anticipate running into the kitchen table. Thus, most people would probably anticipate a hallway, carpet, walls, and maybe a distant railing for the stairs in their spatial 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.

5.2.2 Modularity. Kumaran and Maguire (2009) argue in opposition to a view that sees memory and perception as discrete neural modules. This perspective is present in the way the hippocampus acts, in a top-down fashion, on neural firing in the visual cortex to create the experience of anticipatory extension (Chadwick, Mullally, & Maguire, 2013). The result is that perception is motivated by neural processing and memory as well as by environmental input.

Research by Howard, Kumaran, Ólafsdóttir and Spiers (2011) extended the initial view to better differentiate the CA1 and CA3 subregions of the hippocampus while further motivating the integrative capacity that would be later explored by Chadwick, Mullally, and Maguire

(2013). In this view, the CA1 receives input from both external sensation through the entorhinal cortex and input from memory through the CA3 subregion. Input and output from perception informs memory and vice versa. In this division, 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).

Coherencer follows similar divisions. Its buffer and seeding process (the top- n section in Figure 2) is close in functionality to the CA3 associative subregion in that both are the interface with mnemonic recall and the associated memory processes. However, though processes are distinct across the two systems, they intimately inform one another: Coherencer cannot return anything without the initial structure (i.e., the most highly co-occurring objects with a given query) that is returned from memory. The thresholding mechanism, which mirrors CA1 by detecting whether or not the set matches its sense of coherence, then closely informs when new objects need to be selected from memory. In addition, this entire process, from input to CA3 to CA1, encompasses the traditional psychological view of recall (i.e., not mnemonic recall).

5.3 Differences between Coherencer and SCT

5.3.1 Coherence or Spatial Navigation. Mullally, Intraub, and Maguire (2012) state explicitly that patients with damage to the hippocampus, that were unable to experience boundary extension, 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.”

What they were unable to do was determine where they 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 dissociation between Coherencer and SCT. However, I argue that this is not the case.

First, when the authors are talking about spatial integration, what they mean is a scene that possesses a spatial arrangement of the objects that is appropriate to the real world (Maguire, personal communication). Thus, 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 distinct from Coherencer's concept of coherence, but it is related. As previously stated, spatial properties assume co-occurrence. What is unclear is whether the "additional objects" that the patients could generate were coherent. Given that the patients could not spatially integrate, the association between spatial integration and coherence through co-occurrence, and Coherencer's overall description of the process, one would predict that the additional objects were incoherent (or had a very low coherence score according to Coherencer). Nevertheless, Mullally, Intraub, and Maguire claim that there is some association.

As I will discuss in detail later, it is my suspicion that the collection of these objects *as a whole* would be incoherent: the associations returned are actually the top-*n* co-occurring objects initially queried by Coherencer. In this way, they would appear qualitatively associated in that they combine well with the original query, while still being of a low coherence as per the model. Problematically, the set of responses given by patients in the assessment of coherence was not reported by Mullally, Intraube, and Maguire (2012). Thus, one can only speculate at present.

5.4 Coherencer's Contributions to SCT

5.4.1 Functional loops. In the context of the CA1 and CA3 divisions of the hippocampus discussed above, recognized mismatch when generating a scene would encourage the processing to stay bound in the hippocampal loop. Potentially this would be propagated into the visual cortex as per the top-down features previously discussed. Depending on the complexity of the

interaction, Coherencer predicts that the sensory input processed by the entorhinal cortex might not *always* be external to the organism: top-down propagations might create a larger cortical loop of hippocampal anticipations to the visual cortex to further hippocampal anticipations. The length of the cortical loop, shorter for the hippocampally localized and longer from the hippocampus to visual cortex, would then give a reasonable account of the difference between conscious and unconscious processing. The shorter loop would occur too quickly for the organism to be aware of it while the longer loop, especially in its interaction with the visual cortex, would give the right conditions for conscious experience.

5.4.2 A serial process. Coherencer predicts that the structure of the scene generation in the hippocampus would functionally be a serial process like the object swapping in the model. Computationally, a parallel description of the combinatorial transitions leads to problems in achieving optimality (even approximately, as per Coherencer). For example, if one imagines each possible object swap as a step in one direction (e.g., forward, back, left, right), then swapping two objects is the same as taking two steps in two of these directions. In a square grid where each new direction available upon arriving at a new position is checked, taking two steps means that two directions in the intermediary step will not be checked. To summarize, parallel processing is problematic when a particular step is in the right direction or not based on what *other* steps are taken. This is not to say that it would be impossible to run multiple of these serial processes in parallel or to perform the individual checks at each step in parallel. However, both of these suggestions are also consistent with Coherencer.

Because the brain operates in parallel at low levels, proposing a serial process on a parallel architecture requires explanation. The transitions between neural structures like the CA1 and CA3 are at least functionally serial in as much as the processing of one step in the loop

informs processing in later steps in the loop (see Amaral & Witter, 1989; Amaral, 1993), much like the combinatorial process just described. This offers one possible explanation for the serial process. Additionally, Coherencer's 'objects' are in fact composed of multiple parts and features in the real world (e.g., a nose is a component of a face) all of which are represented with patterns of neural firing. Thus, even if there are exceptions to the serial structure, it is entirely possible that the tighter loops (i.e., effectively faster transitions that produce greater parallelism) are varying minor changes in the object structure or even more minor changes in the neuronal firing. The apparent parallelism could also function as a holographic structure of hierarchically serial processing, where larger loops are tied to larger objects or scene parts. Though this is entirely speculative, the point is that, at the coarse level of functionality endorsed by Coherencer, serial processing can variously emerge on even an entirely parallel architecture like the brain.

5.4.3 The coherence of spatial integration. The easiest immediate test of Coherencer's extension of the work in SCT requires the verbal data from Mullally, Intraube, and Maguire (2012) when they asked patients and controls to give examples of objects that 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 should be low. That is, I predict that patients with hippocampal damage will produce incoherent images in their imagination. However, should the results show that the patients can generate combinations of objects that are of a high coherence, it would suggest that Coherencer should be more precisely teased apart. 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.

This sequencing is already present in the basic structure of Coherencer. As previously stated, the first step performed by the model is an initial seeding of its buffer with the most

probable co-occurring objects with the trigger or query. This assumes that, by some functionality present in the mnemonic recall system, the initial triggering in the environment is somehow associated relatively directly to the initial selection.

What one would expect on the basis of this model, then, is that patients with hippocampal damage would use the objects recalled in this seeding process to inform their responses. 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 the initial seeding as a consequence of skipping the contextual refinement that Coherencer models.

5.5 Discussion

5.5.1 Other theories. Maguire and Mullally (2013) expand the differentiation proposed by Hassabis and Maguire (2007) to directly exclude *purely* memory-based accounts of hippocampal function like the relational theory and the constructive episodic simulation hypothesis (see, for example, Schacter, Addis, Hassabis, Martin, Spreng, & Szpunar, 2012). In these views, the hippocampus either binds the disparate elements of a scene into a single unit or extracts episodic details from memory and combines into a fictitious simulation, respectively. In effect, the hippocampus would function as a stitching algorithm like some computational image or video blending techniques, rather than an anticipatory structure with its own kind of intrinsic reasoning. Maguire and Mullally (2013) argue that, though the hippocampus might perform these roles, they cannot be the whole story.

Coherencer’s use of co-occurrence probabilities is also consistent with both the relational and constructive episodic simulation hypotheses, in that co-occurrence probabilities imply a limited sense of selection and integration (e.g., a non-zero co-occurrence value between two

objects is required for selection). As a consequence, if empirical research demonstrates that these elements are in fact included in the processes of the hippocampus, Coherencer is consistent with those theories as well. If they are relegated to an earlier step in the process, Coherencer would still be consistent with that prior step.

5.5.2 Summary. Recent research in neuroscience has suggested an alternate theory of hippocampal function that offers a new avenue for the integration of cognitive, neuroscientific, and computational research. The authors of said theory, SCT, propose that this integration is possible but it requires as yet unachievably complex computational models of the brain. I offer an alternate avenue on the basis of a functional model that has many parallels with the theory and a number of contributions.

I chose to focus on three contributions. The first tentatively suggested that functional loops of different scales in the hippocampus and related areas could give predictions on where and how the scene construction process might be conscious or unconscious. The second contribution argued that a functionally serial description of the scene construction process was consistent with the requirements of optimization and the underlying neuroscience. Finally, I offered an experimental prediction for the output generated by patients with damage to their hippocampus when asked to describe objects that would be present just outside a local scene: the objects will be locally coherent with the query but incoherent when considered as a group. Implications of various outcomes in this experiment were also discussed.

5.6 Conclusions

A number of contributions have been made in this section to the study of contextual coherence in general and in the visual imagination. To begin, I can expand the conclusions from the previous section (4.7). First, I can conjecture that, if the hippocampus does operate similar to Coherencer

as I have preliminarily shown in this section, it can potentially be used as a biological exemplar of optimization mechanisms in the brain. This would then naturally extend to decompression procedures for lossy forms of compression through Coherencer's association with the domain. Second, there is tentative support for the notion that Coherencer is a good empirical, computational, and *neuroscientific* benchmark for the visual imagination and contextual coherence across *memory* representations, in as much as memory is the primary method for storing representations in a neural context. Third, I can tentatively conclude that the quantitative evaluation of contextual coherence has preliminary neuroscientific support since it is at least consistent with the (re)constructive memory hypothesis. Fourth, I can tentatively conjecture that viewing coherence generally in terms of compression and decompression has neuroscientific support since the hippocampus is consistent with this framework and Coherencer, as a preliminary, information processing model of the hippocampus, can capture context information. Fifth, I can conjecture that research into the hippocampus, in as much as it is performing a decompression process, might offer valuable insights into research on decompression techniques and, consequently, domains using lossy reduction. Sixth, since the types of compression contribute to the overall optimality of the system, it is reasonable to conjecture that research on compression in the brain would offer valuable insights to a full description of the fully functioning hippocampus. Simultaneously, the comparison of information processing descriptions of the hippocampus across compression techniques should impart valuable insights into its functioning in as much as it allows one to differentiate the different elements of these processes. Seventh, SCT is very suggestive that the hippocampus is one of the primary areas associated with generative cognition, construed as a decompression mechanism that operates across many domains.

In terms of Section 3.7, I can add the following conclusions. First, I can conjecture that variations in the functionality of the hippocampus might differentiate the serial-parallel divide described by Thagard (2000) *and* the associated bounded-optimal divide. Thus, humans might improve their optimality by a greater parallelization of the underlying, functionally serial architecture of the hippocampus. The possibility of child and adult neurogenesis in the hippocampus is suggestive in this regard (Andersen, Morris, Amaral, Bliss, & O'Keefe, 2006). Research focusing on developmental and evolutionary changes in parallelization would be suggestive in this regard, both for the theory and for the achievement of computational models that better approach optimality.

7. General Discussion and Conclusions

7.1 Introduction

This work has focused on a number of contributions to the understanding of cognition, computation, and neuroscience. The overarching goal has been to analyze and discuss the intricacies of contextual coherence in the visual imagination. Towards this effect, I initially framed the work in terms of a computational model, SOILIE, of the visual imagination (Section 2). I then proposed an augmentation, Coherencer, of one sub-system of this model, the top- n approach, and demonstrated that it significantly improved the coherence of 5-label sets using a novel, quantitative evaluation.¹¹

In Section 3, I explored the work of Paul Thagard (2000) on contextual coherence in humans in general. I then implemented his proposed model and compared it with Coherencer. Surprisingly, Coherencer still offered a significant improvement over Thagard's model despite the fact that its incremental approach should have been sub-optimal.

In Section 4, I expanded the explanation of the coherence problem in terms of computational compression and decompression. I then compared Coherencer and top- n across two compression techniques: the original co-occurrence probability representation and a holographic vector representation. It was demonstrated that the better decompression approach, Coherencer, was consistently better across both compression representations; thus, I further motivated the need to study models like Coherencer, decompression algorithms, and the task design originally instantiated in Section 2.

¹¹ It should be noted that this and all future conclusions and conjectures in this section should and are intended to be constrained by the fact that the research is specific to a single dataset: the Peekaboom database. Generalizations beyond this dataset, even when implied, *necessarily* require future research. The intent is to merely point to potentially productive research vectors rather than make any claim to absolute truth.

In Section 5, I showed that Coherencer is well grounded in the underlying neuroscience of the hippocampus, specifically Scene Construction Theory (SCT). After demonstrating the parallels and a few differences, I showed that the model could make useful experimental predictions for the underlying neuroscience and that future conclusions from within neuroscience could further inform the model.

In sum, I have established that contextual coherence in the visual imagination is a fascinating, highly interdisciplinary domain that informs as much as it is informed by the related domains of general coherence, computation, and neuroscience. I will now address the contributions to each of the sub-components of this project individually. However, it will be worth addressing some potential problems of the work first.

7.2 Problems

7.2.1 Parameter search and validation. One issue is a consequence of how I selected the thresholds for Coherencer, and Thagard's model. For Coherencer on the co-occurrence representation the threshold was determined by running the given model 10 *on every label in the database* for each threshold starting at 0.01 and going to 0.6 in increments of 0.01 (see Figure 5). Note that for the minimum threshold (0) and maximum threshold (1.0, every label occurs with every other label in every image) the model had similar scores (4842 and 7728.3, respectively) to 0.04 and 0.6 on the chart. On the holographic representation, thresholds were tested from 0.1 to 1.0 in increments of 0.1 (see Figure 6).

Figure 5: Threshold search for Coherencer on co-occurrence representation

Figure 6: Threshold search for Coherence on holographic representation

For Thagard's model, the same thing was performed 5 times but starting with 0.001 and going to 0.05 in increments of 0.001 (see Figure 7). Note that for the minimum threshold (0) and maximum threshold (1.0, every label occurs with every other label in every image) the model had far worse scores (2563.1 and 2531.2, respectively) to 0.001 and 0.05 on the chart. The scores across all runs for each condition were then averaged and the lowest high scoring threshold was

used by each model (0.5, 0.9, and 0.005, respectively). Initialization, max and min activation, and decay used the same parameters as described by Thagard while the max iteration limit was simply an upper threshold that rarely was surpassed. For all 40 runs in Section 3.5, the same 4 labels failed to pass the threshold: “actress,” “braid,” “gay,” and “whiskey.”

Figure 7: Thagard's model threshold search

Though searching for parameters using the entire dataset is non-standard in the modeling literature as it biases the model to the given dataset, I believe there is justification for this approach. Context is an inherently relative affair (i.e., there is no absolute, universal context to which all people should appeal) and, in many ways, it is the very *embodiment of bias*. That is, to be contextual is to be biased to that context over other contexts. I suggest that the model learns this threshold over the course of its experiences with its environment. Consequently, this threshold will be unique for each individual and dataset. For these reasons I ignored this modeling heuristic.

7.2.2 Overfitting and prediction. Throughout this work I have chosen to model contextual coherence in the visual imagination in terms of compression and decompression. In this

framework, I abandoned the common test-train design that is common in the machine learning literature and ‘trained’ the model on the entire dataset. This makes sense because it would be incomprehensible to compress half or some portion of the data and expect to get the entire dataset back upon decompression. However, some might argue that this biases the model, or at least the evaluation, to the current problem. I counter that this is the entire point of decompression: when one decompresses a given dataset from its compressed state, one does not want to produce some universal dataset; one wants *only* the relevant information that is specific to this context. This still might be unconvincing.

Hawkins (2004) describes the problem of overfitting in detail in relation to regression problems, broadly defined: problems that explore the relationship between a dependent and independent set of variables. In the current work, there are two ways of setting up this relation. In terms of the statistical analyses, one can see the dependent variable as the number of label sets that are or are not contained in one of the images from the Peekaboom database: the operationalization of contextual coherence. I then mainly looked at the models (e.g., Coherencer, top- n) as the independent variable, but I also looked at the type of compression in Section 4. Hawkins describes this regression setup as “effect quantification” where there are no future unknowns that want to be predicted, but one wants to better understand how the dependent variable’s distribution is established by the independent variables. The alternate setup has the same dependent variable but views the independent variable as the co-occurrence of labels. This independent variable is distinct from co-occurrence probabilities in that it does not view them as a form of compression. Instead, co-occurrence is just a feature in the environment that might or might not accurately predict label set inclusion (i.e., contextual coherence) when considered in some linear or non-linear relation to it. Hawkins describes this setup as a “prediction problem”

where one wants to anticipate future cases where they know the co-occurrence of labels but do not know their context. These two setups are separate, but related.

Hawkins then outlines two main reasons for overfitting: an overly flexible model (e.g., a neural network's non-linear modeling is more flexible than straight linear regression) and the inclusion of irrelevant independent variables. For the first setup, I used a simple chi-square test on all evaluations except one where I used a simple logistic regression model with two features and their interaction. Neither of these is overly complex and, since each feature had a statistically significant relation to the dependent variable, it would be strange to say they could possibly be irrelevant. In the second setup, I used the top- n , the mean of a matrix, and the mean of a matrix added to the total number of co-occurring terms with various sampling approaches (e.g., direct selection, random walk through co-occurring terms without replacement, parallel random walk through co-occurring terms with replacement) as the model. Again, these are relatively simple and, since I only used a single feature, it would be strange to say it is irrelevant.

Hawkins goes on to say that the main problems with overfitting are wasted resources, poor decisions, additional random variation, and lower portability. The first three are mainly the cause of unnecessary variable inclusion, which is not a problem in either setup. For the last one, Hawkins explicitly states that one predictor models are very portable and I can probably reasonably extend this to two predictor models. Even the sampling approaches are relatively easy to implement. In effect, neither setup is liable to have a problem with overfitting; thus, the absence of the train-test design is irrelevant.

Finally, if this is still not convincing, I can directly exclude the predictive setup from the current analysis. The result is that the conclusions make no claims on the generation of novel, but coherent, contexts (i.e., a context that has never been experienced or is not in the database and

yet is still coherent). However, it is unclear exactly what this would mean as it is unclear whether humans ever generate truly novel contexts (see Cockbain, Vertolli, & Davies, 2014); or, if contextual coherence literally means *intelligible relative to what one has already experienced*. Given this interpretation, one would predict that one *must* have some experiences of a given context in order for it to be even remotely coherent and novelty might be relatively proportional to how small the set of experiences for that context actually is. Alternatively, novelty might just be a particular balance on a coherence-incoherence spectrum that manages to define a new context. These claims follow from the given setup, but I propose them tentatively in order to indicate the bounds of the current research.

7.3 Contributions

7.3.1 Visual imagination. A number of contributions have been made to the study of contextual coherence in the visual imagination. I have provided a useful description of the problem: compression representations (e.g., co-occurrence probabilities) that do not explicitly represent context relations—they are *lossy* in this respect—will need more complex decompression techniques. Optimization techniques like Coherencer’s incremental approach were shown to be successful in this domain. SCT’s view of the hippocampus adds preliminary biological support to this view.

I can also conclude that at least some domains in the visual imagination are of a simple representation, low dimensionality, and high combinatoric load. I tentatively conjectured that some systems might capitalize upon greater parallelization either through learning or neurogenesis. This suggests an interesting continuum of bounded contextual coherence processes to optimal as a consequence of greater parallelization.

I have proposed a new, quantitative method for assessing contextual coherence in the visual imagination: awarding points for generated label sets that occur in an image in the uncompressed database. The key here is that, if context is lost in the process of compression, its deduction is a measurable achievement. This evaluation is consistent with the constructive memory hypothesis and the underlying neuroscience described by SCT. One caveat is that the evaluation procedure requires known, approximately optimal samples in an uncompressed form of the underlying visual coherence. I have proposed that real world pictures minimally satisfy this constraint.

There is support for the notion that Coherencer is an improved system for deducing contextual coherence in the visual imagination from the top- n model. This is true across at least two different compression representations: co-occurrence probabilities and holographic symbolic architectures. I conjectured that one can then extend this conclusion to view memory in terms of this compression and add the model as a benchmark for the underlying neuroscience, as well. Future research is necessary to determine just how good it would be in this respect. Consequently, future computational models of the visual imagination can use Coherencer as an empirical and computational benchmark (i.e., there is a preliminary measure that can now be used to ground future empirical research in this domain).

7.3.2 Contextual coherence. A number of contributions have been made to the study of contextual coherence in general. First, I used an analogy with compression and decompression to explain coherence. In this view, potentially relevant cues for the designation of a given context can be understood in terms of lost information (i.e., the actual-potential dyad is transposed to what is kept-lost in a given compression process). The characterization of compression as the reduction and abstraction of irrelevant details of a stimulus to those that are relevant for the

survival of the organism lends tentative support for this position. The work on holographic symbolic architectures adds further support by demonstrating the loss of context and its deduction/decompression across compression techniques. Neuroscientific support was also added when Coherencer was shown to functionally model the hippocampus and successfully deduce lost context information.

I can conclude that Coherencer can be used as an empirical and computational benchmark for research on coherence. I can also conclude that at least some contexts are of a simple representation, low dimensionality, and high combinatoric load. I tentatively conjectured that simpler systems should better fit this definition while more complex systems might achieve this complexity through greater parallelization of various forms.

I can conjecture that the quantitative evaluation that was used for contextual coherence in the visual imagination should generalize broadly. The basic setup should work as long as a given modality can be examined in terms of multiple elements (e.g., labels, words, images, scents) which need to be selected in terms of one another and from multiple potential contexts. Fill-in-the-blank sentences were given as an example for textual context relations. This is consistent with the underlying neuroscience since the hippocampus operates on and integrates all sensory modalities.

Finally, I conjectured with reference to research on human creativity (Cockbain, Vertolli, & Davies, 2014) that contextual coherence could be necessarily restricted to that which is already experienced. I then offered the tentative prediction that novelty might merely be an indication of a context that has been experienced less or, alternatively, novelty might be incoherence of a given degree such that it defines a new context.

7.3.3 Decompression. A number of contributions have been made to the study of decompression. In order to situate these claims, I first defined compression as the reduction and abstraction of irrelevant details of a stimulus to those that are relevant for the persistence of a system. Given this position, decompression is the deduction of some set of these details in order to produce output that is not explicitly represented due to loss or for other reasons. Through the notion of generative cognition, I can conjecture that decompression techniques might be implicated in the full range of human faculties that handle compressed data, largely in terms of memory, in as much as memory is a compressed representation of the world.

There is tentative support for the claim that better decompression techniques will improve optimality across compression representations for data that is not explicitly represented. This led to the tentative normative conclusion that research in domains that require this type of compression need to explore the implications of the decompression techniques being used. Future research is necessary to further support this conclusion.

I supported the claim that optimization techniques like Coherencer perform well as decompression procedures for this type of compression. Specifically, I showed that decompression can be used to generate contextual coherence. Scene construction theory coupled with the Coherencer model added preliminary biological support for these conclusions by viewing the hippocampus as a decompression mechanism. Research focusing on developmental and evolutionary changes in the hippocampus would, consequently, be informative in this domain.

There is evidence for the idea that the parallel-incremental distinction proposed by Thagard does not guarantee the quality of the decompression procedure. I can also tentatively suggest that domains with simple representations, low dimensionality, and high combinatoric

load might be better suited to incremental decompression procedures. Situations with high complexity, high dimensionality, and lower combinatoric load could extend from these lower systems. Finally, I conjectured that research into the hippocampus might offer valuable insights into research on decompression techniques and, consequently, domains using lossy reduction.

7.3.4 Generative cognition. A number of contributions have been made to the study of generative cognition. There is support for the notion that generative cognition can be construed as a form of decompression, possibly optimization. The work on Scene Construction Theory supports this conclusion in both the imagination and memory, and associates it with the hippocampus. Thus, research into the hippocampus might offer valuable insights into this domain. Coherencer was shown to be a working model of these processes at least in terms of contextual coherence. Similarly, in as much as generative cognition necessitates the creation of a context, I take contextual coherence to be of central interest for research on generative cognition.

7.3.5 Holographic memory. A number of contributions have been made to the study of holographic memory. First, given how poorly the holographic representation performed on the contextual coherence task, I can tentatively conclude that this task is either a slightly different domain or more challenging than the evaluation techniques used in the associated literatures on holographic symbolic architectures. I then conjectured that variations of this evaluation could be used as a more stringent testing mechanism in related domains (e.g., the holographic vector research on texts). Future research is necessary to demonstrate this.

The low success rate of the holographic vector representation compared to co-occurrence probabilities across both decompression techniques is a concerning finding. At minimum it preliminarily suggests a domain where holographic vectors do poorly. It also tentatively suggests that the current tests for the representation are potentially too easy.

Finally, the drastic improvement that Coherencer gave to the holographic vector representation tentatively suggests that the study of decompression techniques in this literature is important. It also suggests that Coherencer could be used to augment the already existing approaches in this literature.

7.3.6 SCT and the hippocampus. A number of contributions have been made to the study of Scene Construction Theory (SCT). First, I can conjecture that research into the hippocampus might offer valuable insights into research on decompression techniques and, consequently, domains using lossy reduction.

Second, since there is preliminary evidence that the types of compression contribute to the overall optimality of the system, research on compression in the brain is important for capturing an accurate description of the full functionality of the hippocampus. Similarly, the comparison of information processing descriptions of the hippocampus across possible compression representations would likely impart valuable insights into the underlying hippocampal functionality especially as better descriptions of the memory compression are achieved.

There is support for the notion that variations in the functionality of the hippocampus might differentiate the serial-parallel divide described by Thagard (2000) *and* the associated bounded-optimal divide. Humans could improve their optimality by a greater parallelization of the underlying, functionally serial architecture of the hippocampus. The possibility of child and adult neurogenesis in the hippocampus is suggestive in this regard (Andersen, Morris, Amaral, Bliss, & O'Keefe, 2006). Research focusing on developmental and evolutionary changes in parallelization would be suggestive in this regard, both for the theory and for the achievement of computational models that better approach optimality in these domains.

I proposed that patients with damage to their hippocampus should only perform Coherencer's top- n seeding process when tasked to recall other objects in the surrounding environment outside of a visual scene. The average coherence produced by this process should be quite low, in domains with multiple contexts.

7.4 Miscellaneous Assessments

A few minor extensions to the current research were briefly explored without full, quantitative and statistical evaluation. I will mention them here in brief. First, I tried to improve Coherencer by adding a second threshold for the standard deviation of the rows and columns of the co-occurrence probability matrix. I found that this added only minor improvements (less than 10 additional successes over most runs). Another avenue that was explored was adding a seeding process to Thagard's model. In this approach, nodes of the top-4 co-occurring labels with the query were started with a higher activation: 0.5 instead of 0.01. This also offered little improvement.

7.5 Future Research

There are three main avenues by which I am extending this research. The first extends this research into the field of generative cognition broadly construed in terms of 3D environments. This research is focused on gathering further insights from neuroscientific research on the hippocampus, integrating spatial parameters like distance and size into the current processes, finding better compression techniques like clustering metrics, and incorporating a chunking mechanism in order to deal with larger numbers of objects. Parallel research on object and pattern recognition in visual domains is also being examined to improve and automate the labelling process.

The second stream is focused on better integrating the current research with the holographic symbolic architecture literature. This research is focused on using textual corpus data to inform visual coherence relations, testing the effects of Coherencer on holographic vectors with exclusively textual corpus data, and designing and evaluating more sophisticated decompression techniques.

The third and final stream is exploring formal approaches to describing the task and models (e.g., Bayesian theory, research on information retrieval, and type theory). I am also examining Bayesian data analysis in hopes that it might offer a more robust evaluation framework for future research.

Appendices

Appendix A Properties of the Peekaboom Database

Some very basic properties of the filtered database follow. An important caveat is that datasets that are very different from the one described herein are likely to result in very different results. Which properties are the most salient is as yet unclear, but they are included here in order to be comprehensive for future insights that are not yet discovered.

First, Figure 5 lists the distribution of labels with a particular number of images containing them. The greater majority of labels have only one image associated with them ($n = 3,463$). The distribution of the number of co-occurring terms with these labels is shown in Figure 6. The maximum number of images associated with a label ('man') is 5,277. The mean is 17.48, median is 2.00 and standard deviation is 114.27.

Figure 8: Histogram¹² of the quantity of labels with a given number of images containing it.

Figure 9: Histogram of the quantity of labels with one image with a given number of co-occurring labels.

¹² In all histograms in the horizontal axis, the bins for the distribution are designed such that the left edge is less than or equal to all values in the bin, which are less than the right edge (i.e., for label x in the first histogram in the first bin, $1.0 \leq x < 2.0$ so the values range from 1.0 to 1.999...).

The next basic property is the distribution of co-occurring terms for each label (see Figure 11). The maximum number of co-occurring terms associated with a label ('man') is 2,975. The mean is 38.50, median is 11.00 and standard deviation is 104.22.

Figure 10: Histogram of the number of labels with a given number of co-occurring labels.

Using the NetworkX library (Hagberg, Schult, & Swart, 2008) a number of database properties were found. The diameter of the database was found to be 5. Thus, starting at any given label one could get to any other given label in a maximum of 5 steps through co-occurring nodes. The radius of the database was found to be 3. Thus, for any given label, the most distant label in the database was at minimum 3 steps through co-occurring nodes.

The periphery of the database, labels whose eccentricity (i.e., maximum graph distance to any label) is equal to the diameter (i.e., 5), included 46 labels (0.55%). The center of the graph, labels with eccentricity equal to the radius (i.e., 3), included 1,092 labels (13.04%). And thus, all remaining 7,234 labels (86.41%) had an eccentricity of 4.

Appendix B Formal Definition

Zhang and Zhou (2013) describe the multi-label task as follows. Let X denote the input space (e.g., images, documents) and Y denote the label space of all possible labels. The standard task is to learn a function $h(\cdot)$ that takes as input some member $\mathbf{x}_i$ from the input space and returns some combination of labels from the label space as output (i.e., $h : X \rightarrow 2^Y$). This function is learned from the multi-label training set D , where all ε training examples are described in terms of their input features and corresponding label sets (i.e., $D = \{(\mathbf{x}_i, Y_i) | 1 \leq i \leq \varepsilon\}$). Note that $\mathbf{x}_i$ is a d -dimensional feature vector $(x_{i1}, x_{i2}, \dots, x_{id})$, where each feature corresponds to a single dimension in the input space. In the standard classification task one wants to find the function $h(\cdot)$, called the classifier, in order to predict the correct labels that go with an as yet unseen input $\mathbf{x}$.

By contrast, the generative task is to find a function $g(\cdot)$ that is the inverse of $h(\cdot)$ (i.e., $g : 2^Y \rightarrow X$). This means that it takes as input a label or set of labels and outputs one of the original input instances (e.g., image, document). I propose to preliminarily achieve this through the parallel task of finding a set of labels Y_μ , where μ indexes the current iteration of the algorithm, that extend the input y and together indicate some instance $\mathbf{x}_i$. One can think of the generation task as finding a function $g(\cdot)$, called the generator, that finds a subset of labels that would be picked by an accurate classifier $h(\cdot)$ for *some* instance $\mathbf{x}_i$ in the input space X .

Formally, this means

$$\exists \mathbf{x}_i. h(\mathbf{x}_i) \supseteq g(y)$$

Since the manual labeling of the Peekaboom database allows one to map each label y to the subset of features that indicated it (e.g., for the label ‘dog,’ a given collection of pixels that look

like a dog), the set of labels Y_μ is equal to the collection of all these feature subsets and approximately equal to the original input instances $\mathbf{x}_i$ that contain them. Thus, $g(\cdot)$ meets the requirements of the given task (i.e., $g : 2^Y \rightarrow Y$ and $Y_\mu \sim \mathbf{x}_i$ then $g : 2^Y \rightrightarrows X$).

Models of both the generative and classifier tasks return a real-valued function $f(\cdot, \cdot)$ that takes an instance-label pair $(\mathbf{x}, y)$ as input and outputs a number denoting the confidence that the label is accurate for that instance (i.e., $f : X \times Y \rightarrow \mathbb{R}$). However, in the generative case, one can assess the confidence of a label y relative to the current potential set of labels Y_μ . If one thinks of Y_μ as a hypothetical instance $\mathbf{x}$, then a modified version of $f(\cdot, \cdot)$'s input, or $(\mathbf{x} \cong Y_\mu, y)$, is achieved for the model of the generator. For both the generator and the classifier tasks, $f(\cdot, \cdot)$ should output a larger confidence value on a relevant label y' than an irrelevant label y'' for a given instance or hypothetical instance $\mathbf{x}$, or $f(\mathbf{x}, y') > f(\mathbf{x}, y'')$ (Zhang and Zhou, 2013). The multi-label classifier $h(\cdot)$ and generator $g(\cdot)$ can then be derived from $f(\cdot, \cdot)$ by incorporating a thresholding function that determines how large the confidence needs to be for a label to be considered accurate for a given instance (i.e., $t : X \rightarrow \mathbb{R}$). The output, $h(\mathbf{x})$ and $g(y)$, is then composed by assessing for a given instance or hypothetical instance $\mathbf{x}$ whether each possible label y passes the threshold given by $t(\mathbf{x})$. Formally, this means

$$h(\mathbf{x}) = g(y) = \{y | f(\mathbf{x}, y) > t(\mathbf{x}), y \in Y\}$$

In effect, both functions use $t(\cdot)$ to dichotomize Y into relevant and irrelevant label sets (Zhang and Zhou, 2013). However, the generative $t(\cdot)$ is slightly more complex as the hypothetical $\mathbf{x}$ changes with each iteration of the algorithm.

Using the manual labelling procedure of the Peekaboom database as $h(\cdot)$ combined with the co-occurrence probabilities of each pair of labels, it is possible to construct the compressed

representation $CP = \{(a, b, P(a|b)) | a, b \in Y, a \neq b\}$ that is used by both the top- n and Coherencer models. From here it is possible to formally define both models.

The top- n model can be formalized as follows. The function $f(\cdot, \cdot)$ just returns the co-occurrence probability of each y given q or $f(Y_\mu, y) = P(y|q)$. In this case $Y_\mu = Q = \{y | P(y|q) > 0\}$. The thresholding function $t(\cdot)$ simply sets itself to the co-occurrence probability of the n^{th} ranked label or $t(Y_\mu) = P(\text{rank}_n(Y_\mu) | q)$, where $\text{rank}_n(\cdot)$ returns said label.

Coherencer can be formalized as follows. First, the label set Y_1 of the initial hypothetical instance $\mathbf{x}_1$ in $f(\cdot, \cdot)$ is defined as the top-4 labels with the highest conditional probability with the query (q) or $Y_1 = \{z_k | \forall y \in Y. P(y|q) \leq P(z_k|q) \leq P(z_j|q), 1 \leq j < k \leq 4\}$. The function $f(\cdot, \cdot)$ acts on the subset of CP , called K_μ , that contains the elements in the current Y_μ and their corresponding conditional probabilities (i.e., $K_\mu = \{(a, b, P(a|b)) \in CP | a, b \in Y_\mu \cup \{q\}\}$). Specifically, the function $f(\cdot, \cdot)$ sums over the conditional probabilities of the subset of triples in K_μ that contain y , $K_{\mu y} = \{(a, b, P(a|b)) \in K | (a = y) \vee (b = y)\}$, or

$$f(Y_\mu, y) = \sum_{u \in K_{\mu y}} P_u$$

The function $t(\cdot)$ evaluates the current total context K_μ and, if it passes a threshold (λ), outputs 0 allowing the algorithm to return $Y_\mu \cup \{q\}$ as a valid set of labels for the generated instance $\mathbf{x}_\mu$. Otherwise, it outputs the minimum $f(\cdot, \cdot)$ value in Y_μ , effectively discarding the associated label. I can express this formally as

$$t(Y_\mu) = 0, \text{ if } \sum_{v \in K_\mu} P_v > \lambda$$

$$t(Y_\mu) = \text{argmin}(f(\mu, y)), \forall y \in Y_\mu, \text{ if } \sum_{v \in K_\mu} P_v < \lambda$$

where argmin returns the lowest $f(\cdot, \cdot)$ value.

```

Y = Coherencer(CP, q)

1. Initialize  $Q, R, C, Y_1$ 
2. For  $\mu = 1$  to  $|Q| - |Y_1|$  do
3.   Set  $K_\mu, K_{\mu y}$ 
4.    $Y_{\mu+1} = \{\}$ 
5.   If  $\sum_{v \in K_\mu} P_v > \lambda$  do
6.      $t(Y_\mu) = 0$ 
7.   Else
8.      $t(Y_\mu) = \min(f(Y_\mu, y)), \forall y \in Y_\mu$ 
9.   For  $y \in Y_\mu$  do
10.    If  $f(Y_\mu, y) > t(Y_\mu)$  do
11.       $Y_{\mu+1} = Y_{\mu+1} \cup \{y\}$ 
12.    If  $|Y_{\mu+1}| = 5$  do
13.      Return  $Y_{\mu+1} \cup \{q\}$ 
14.    Else
15.       $R \cup (Y_\mu - Y_{\mu+1})$ 
16.       $C = Q - R$ 
17.      If  $|C| > 0$  do
18.         $Y_{\mu+1} = Y_{\mu+1} \cup \text{rand}(C)$ 
19.         $\mu = \mu + 1$ 
20.      Else
21.        Return  $g'(q) \cup \{q\}$ 

```

Figure 12: Formal pseudocode for Coherencer.

Since $|g(q)| = 5$ is a condition for the termination of the search, a new label y' is randomly selected from $C = Q - R$, where $R = \{y'' | f(Y_\mu, y'') < t(x_\mu), 1 \leq \mu \leq (|Q| - |Y_\mu|)\}$. If at any point $|C| = 0$, the result will be

$$g'(q) = \left\{ y \in Y_\mu \left| \begin{array}{l} \text{argmax}' \left(\sum_{v \in K_\mu} P_v \right), \\ 1 \leq \mu \leq (|Q| - |Y_\mu|) \end{array} \right. \right\}$$

where $\text{argmax}'(\cdot)$ selects the index with the highest corresponding value.

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, 2222–2238.
- Amaral, D. G. (1993). Emerging principles of intrinsic hippocampal organization. *Current opinion in neurobiology*, 3(2), 225-229.
- Amaral, D. G., & Witter, M. P. (1989). The three-dimensional organization of the hippocampal formation: a review of anatomical data. *Neuroscience*, 31(3), 571-591.
- Andersen, P., Morris, R., Amaral, D., Bliss, T., & O'Keefe, J. (Eds.). (2006). *The hippocampus book*. Oxford University Press.
- Attneave, F. (1954). Some informational aspects of visual perception. *Psychological review*, 61(3), 183.
- Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. *Sensory communication*, 217-234.
- 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.), *Proceedings of the 11th International Conference on Cognitive Modeling*, Ottawa: 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.

- Byrne, M., D. (2013). How many times should a stochastic model be run? An approach based on confidence intervals. In R. West & T. Stewart (eds.), *Proceedings of the 12th International Conference on Cognitive Modeling*, Ottawa: Carleton University.
- Chadwick, M. J., Mullally, S. L., & Maguire, E. A. (2012). The hippocampus extrapolates beyond the view in scenes: an fMRI study of boundary extension. *Cortex*, 49(8), 2067-2079.
- 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*.
- Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological review*, 107(2), 261.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1), 87-114.
- Davies, J., Atance, C. & Martin Ordas, G. (2011). A framework and open questions on imagination in adults and children. *Imagination, Cognition, and Personality, Special Issue on Mental Imagery in Children*. 31:1-2, 143-157.
- 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., & Thagard, P. (1997). Waves, particles, and explanatory coherence. *British Journal for the Philosophy of Science*, 48, 1-19.

- Hagberg, A., Swart, P., & S Chult, D. (2008). Exploring network structure, dynamics, and function using NetworkX. In G. Varoquaux, T. Vaught, and J. Millman (eds), *Proceedings of the 7th Python in Science Conference (SciPy2008)*, California: Pasadena.
- 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.
- Hassabis, D., & Maguire, E. A. (2007). Deconstructing episodic memory with construction. *Trends in cognitive sciences*, 11(7), 299-306.
- Hawkins, D. M. (2004). The problem of overfitting. *Journal of chemical information and computer sciences*, 44(1), 1-12.
- 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 *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 525-533). ACM.
- Hullett, K., & Mateas, M. (2009, April). Scenario generation for emergency rescue training games. In *Proceedings of the 4th International Conference on Foundations of Digital Games* (pp. 99-106). ACM.
- Hutter, M. (2005). Universal Artificial Intelligence: Sequential Decisions based on Algorithmic Probability. Springer.
- Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, 114, 1-37.

- 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. (Eds.). (2012). *Handbook of imagination and mental simulation*. Psychology Press.
- 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.
- Rolls, E. T. (2008). *Memory, Attention, and Decision Making: A Unifying Computational Neuroscience Approach*. Oxford University Press, Oxford.
- Rubin, D. C., Schrauf, R. W., & Greenberg, D. L. (2003). Belief and recollection of autobiographical memories. *Memory & Cognition*, 31(6), 887-901.
- Ruderman, D. L., & Bialek, W. (1994). Statistics of natural images: Scaling in the woods. *Physical review letters*, 73(6), 814-817.
- Schacter, D. L., & Addis, D. R. (2007). The cognitive neuroscience of constructive memory: remembering the past and imagining the future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1481), 773-786.
- Schmidhuber, J. (2009). Driven by compression progress: a simple principle explains essential aspects of subjective beauty, novelty, surprise, interestingness, attention, curiosity, creativity, art, science, music, jokes. In G. Pezzulo, M. V. Butz, O. Sigaud, and G.

- Baldassarre (eds.), *Anticipatory Behavior in Adaptive Learning Systems, from Sensorimotor to Higher-level Cognitive Capabilities*, Springer.
- Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological review*, 63(2), 129.
- Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual review of neuroscience*, 24(1), 1193-1216.
- 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.
- 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.
- 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, 435-467.
- Thagard, P. (1991). The dinosaur debate: explanatory coherence and the problem of competing hypotheses. In J. Pollock, and R. Cummins (eds.), *Philosophy and AI: essays at the interface* (pp. 279-300). Cambridge, MIT Press.
- Thagard, P. (1992a). Adversarial problem solving: modeling an opponent using explanatory coherence. *Cognitive Science*, 16, 123-149.
- Thagard, P. (1992b). *Conceptual revolutions*. Princeton, Princeton University Press.
- Thagard, P. (2000). *Coherence in thought and action*. Cambridge, MIT Press.

- Thagard, P., Holyoak, K., Nelson, G., and Gochfeld, D. (1990). Analog retrieval by constraint satisfaction. *Artificial Intelligence*, 46, 259-310.
- Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, 53, 1-12.
- Vertolli, M. O., Breault, V., Ouellet, S., Somers, S., Gagné, J. & Davies, J. (2014). Theoretical assessment of the SOILIE model of the human imagination, *Proceedings of the 36th Annual Conference of the Cognitive Science Society*. Quebec City, QC: Cognitive Science Society.
- Vertolli, M. O. & Davies, J. (2013). Visual imagination in context: Retrieving a coherent set of labels with Coherencer. In R. West & T. Stewart (eds.), *Proceedings of the 12th International Conference on Cognitive Modeling*, Ottawa: Carleton University.
- Vertolli, M. O. & Davies, J. (2014). Coherence in the visual imagination: Local hill search outperforms Thagard's connectionist model, *Proceedings of the 36th Annual Conference of the Cognitive Science Society*. Quebec City, QC: Cognitive Science Society.
- Vertolli, M. O., Kelly, M. A., & Davies, J. (2014). Perception and generation as a compression-decompression dyad, *Proceedings of the 7th Conference on Artificial General Intelligence (AGI)*, Quebec City, QC: AGI.
- Von Ahn, L., & Dabbish, L. (2004). Labeling images with a computer game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 319-326). ACM.
- Von Ahn, L., Liu, R., & Blum, M. (2006). Peekaboom: a game for locating objects in images. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 55-64). ACM.
- Wolff, J. G. (2013). The SP theory of intelligence: an overview. *Information*, 4(3), 283-341.

Zhang, M., & Zhou, Z. (2013). A Review on Multi-Label Learning Algorithms. *IEEE Transactions on Knowledge and Data Engineering*, (99).