Categorical Compositional Cognition

Yaared Al-Mehairi, Bob Coecke, and Martha Lewis[(✉)]

Department of Computer Science, University of Oxford, Oxford, UK
yaared.almehairi@gmail.com, {coecke,marlew}@cs.ox.ac.uk

Abstract. We accommodate the Integrated Connectionist/Symbolic Architecture (ICS) of [32] within the categorical compositional semantics (CatCo) of [13], forming a model of categorical compositional cognition (CatCog). This resolves intrinsic problems with ICS such as the fact that representations inhabit an unbounded space and that sentences with differing tree structures cannot be directly compared. We do so in a way that makes the most of the grammatical structure available, in contrast to strategies like circular convolution. Using the CatCo model also allows us to make use of tools developed for CatCo such as the representation of ambiguity and logical reasoning via density matrices, structural meanings for words such as relative pronouns, and addressing over- and under-extension, all of which are present in cognitive processes. Moreover the CatCog framework is sufficiently flexible to allow for entirely different representations of meaning, such as conceptual spaces. Interestingly, since the CatCo model was largely inspired by categorical quantum mechanics, so is CatCog.

1 Introduction

A key question in artificial intelligence and cognition is how symbolic reasoning can be accomplished with distributional representations. [32] present a view of cognition called the *Integrated Connectionist/Symbolic Architecture* (ICS) that incorporates two levels of formal description: "the continuous, numerical lower-level description of the brain", characterized by a connectionist network, and "the discrete, structural higher-level description of the mind", characterized in terms of symbolic rules. ICS is a hybrid approach to the computational modelling of the mind which uses vectors to represent *roles*, from which symbolic structures may be built, and *fillers*, the objects to be manipulated. These roles and fillers are then combined using the tensor product. However, as argued by [7], the tensor product representations used to codify the isomorphism between connectionist and symbolic representations reveal shortcomings. Firstly, the representational space of a concept grows in size as more elements are added to the compound. Secondly, it is unclear how to compare representations that have differing underlying structures. For example, the concepts *joke* and *funny joke*, have structurally different representations, and it is not obvious how one relates

to the other. The third and last problem is to do with particular implementations of ICS. [15,33] use tensor binding representations to model grammar in the following manner: words such as *funny* or *joke* are tagged with their parts of speech. This tagging takes the following form. Each part of speech is assigned a vector, and each word token is assigned a vector. The outer product of the word vector and the part of speech vector is taken, giving a matrix. Circular convolution [26], consisting of taking the sum over diagonals of the matrix, is then used to reduce the matrix back down to a vector. However, this ignores the fact that parts of speech have different roles and structures.

The problem of unifying symbolic and distributional representations has been addressed in the field of computational linguistics. Distributional semantics [30,34] provides vector meanings for words, but has no clear compositional structure. In contrast, compositional approaches such as that of [23] are able to compute the meanings of phrases, but must take the meanings of words as given. The CatCo model of [13] unifies the distributional theory of meaning in terms of vector space models and the compositional theory of grammatical types. It utilises grammar to derive the meaning of a sentence, represented by a vector, from the word vectors that make up the sentence. This model uses composite spaces without increase in size of the resulting meaning space and allows composite concepts to be directly compared with their constituents of the same type, as well as the meaning of sentences of varying length and structure to be compared. Further, this model explicitly recognises the differing structures of parts of speech, and uses these structures to compute the meaning of the sentence. We use the ideas in CatCo to improve the representation of grammar in connectionist frameworks. At the same time, we reformulate the representations in CatCo so that compositionality can be implemented in a more cognitively realistic setting.

We accommodate ICS within CatCo, forming a new model of categorical compositional cognition (CatCog). The CatCo model was greatly inspired by the categorical semantics for quantum teleportation [6] and nicely matches the template of a quantum-like logic of interaction [8]. Hence, we obtain a model for cognition that draws inspiration from quantum theory.

2   ICS Architecture

In [32], the authors implement symbolic structures within a connectionist architecture. They use vectors and tensor products to represent objects, roles, and structures. Recursive structures such as trees can be represented. The contents of the leaves are encoded in the fillers, and role vectors encode the tree structure. Fillers are bound to roles using the tensor product, and collections of roles and fillers are combined using vector addition.

Recursive Connectionist Realization. Embedding and recursion require the tensor product representation to handle embedded structure, where the filler is itself a complex structure, and not an atomic symbol. The binding $\mathbf{f}/r$ of a filler $\mathbf{f}$ to a role r is realized as a vector $\mathbf{f}/\mathbf{r} = \mathbf{f} \otimes \mathbf{r}$ that is the tensor product of a

vector $\mathbf{f}$ realizing $\mathbf{f}$ with a vector $\mathbf{r}$ realizing r. A sentence $\mathbf{s}$ is represented as a sum of filler/role bindings $\sum_i \mathbf{f}_i \otimes \mathbf{r}_i$, and these can be applied recursively. For example, let $\mathbf{s} = \texttt{[NP VP]}$ be a binary tree with left and right subtrees $\texttt{NP}$ and $\texttt{VP}$. Let $\mathbf{s}, \mathbf{v}_1, \mathbf{v}_2$ be the vectors realizing $\mathbf{s}, \texttt{NP}, \texttt{VP}$. The connectionist realization of $\mathbf{s}$ is:

$$\mathbf{s} = \mathbf{v}_1 \otimes \mathbf{r}_0 + \mathbf{v}_2 \otimes \mathbf{r}_1 \tag{1}$$

If $\texttt{VP}$ is a tree rather than an atomic symbol, it can be expressed in terms of its left and right subtrees $\texttt{VP} = \texttt{[Vt NP]}$. If $\mathbf{v}_3, \mathbf{v}_4$ represent the trees $\texttt{Vt}, \texttt{NP}$, then the structure $\mathbf{s} = \texttt{[NP [Vt NP]]}$ has the following representation:

$$\mathbf{s} = \mathbf{v}_1 \otimes \mathbf{r}_0 + (\mathbf{v}_3 \otimes \mathbf{r}_0 + \mathbf{v}_4 \otimes \mathbf{r}_1) \otimes \mathbf{r}_1 \tag{2}$$
$$= \mathbf{v}_1 \otimes \mathbf{r}_0 + \mathbf{v}_3 \otimes (\mathbf{r}_0 \otimes \mathbf{r}_1) + \mathbf{v}_4 \otimes (\mathbf{r}_1 \otimes \mathbf{r}_1) \tag{3}$$
$$\equiv \mathbf{v}_1 \otimes \mathbf{r}_0 + \mathbf{v}_3 \otimes \mathbf{r}_{01} + \mathbf{v}_4 \otimes \mathbf{r}_{11} \tag{4}$$

The notable feature of this representation is that the vector space in which concepts live must be arbitrarily large, depending on the size of the structure to be represented. Symbols at depth d in a binary tree are realized by $\mathcal{S}_{(d)}$, the FR^d-dimensional vector space formed from vectors of the form $\mathbf{f} \otimes \mathbf{r}_i \otimes \mathbf{r}_j \otimes \cdots \otimes \mathbf{r}_k$ with d role vectors, where F is the dimension of the filler vectors $\mathbf{f}$ and R is the dimension of the individual role vectors $\mathbf{r}_i$. A vector space containing all vectors in $\mathcal{S}_{(d)}$ for all d is:

$$\mathcal{S}^* \equiv \mathcal{S}_{(0)} \oplus \mathcal{S}_{(1)} \oplus \mathcal{S}_{(2)} \oplus \cdots \tag{5}$$

Vectors $\mathbf{s}_{(i)}$ are embedded into this space, meaning that the normal operation of vector addition can be used to combine sentence components.

Symbol Processing. Information is processed in the mind/brain by widely distributed connection patterns (i.e. weight matrices $\mathbb{W}$) that, for central aspects of higher cognition, possess global structure describable through symbolic expressions for recursive functions. [32] show how basic symbol manipulation can be achieved using a distributed system via matrix multiplication. Central aspects of many higher cognitive domains (including language) are realized via recursive processing. Feed-forward networks and recurrent networks provide a mechanism to compute a large class of cognitive functions with recursive structure. In either case, $\mathbb{W}$ is a finite matrix of weights that specifies a particular function (Table 1).

Table 1. Space of descriptions in ICS

Structure	Symbolic	Connectionist
Set	$\mathbf{f}$	$\mathbf{f} \in V_F$
String	$\mathbf{f}_i / r_i$	$\mathbf{f}_i \otimes \mathbf{r}_i$
Tree	$\mathbf{s} = \{\mathbf{f}_i / r_i\}$	$\mathbf{s} = \sum_i \mathbf{f}_i \otimes \mathbf{r}_i \in \mathcal{S}^*$

3   CatCo Semantics

In this section, we summarize the categorical compositional semantics introduced in [13], describing a method for constructing the meanings of sentences from the meanings of words using syntactic structure.

3.1   Pregroup Grammars

Lambek's pregroup grammars [20] are used to describe syntactic structure. This choice of grammar is not essential, and other forms of categorial grammar can be used, as argued in [10]. A pregroup $(P, \leq, \cdot, 1, (-)^l, (-)^r)$ is a partially ordered monoid $(P, \leq, \cdot, 1)$ where each element $p \in P$ has a left adjoint p^l and a right adjoint p^r, such that the following inequalities hold:

$$p^l \cdot p \leq 1 \leq p \cdot p^l \quad \text{and} \quad p \cdot p^r \leq 1 \leq p^r \cdot p \tag{6}$$

The pregroup grammar $\mathsf{Preg}_\mathcal{B}$ over an alphabet $\mathcal{B}$ is freely constructed from the atomic types in $\mathcal{B}$. In what follows, $\mathcal{B} = \{n, s\}$. The type s is used to denote a declarative sentence and n to denote a noun. A transitive verb can then be denoted as $n^r s n^l$. If a string of words and their types reduce to the type s, the sentence is judged grammatical. The sentence *Clowns tell jokes* is typed $n \, (n^r s n^l) \, n$, and can be reduced to s as follows:

$$n \, (n^r s n^l) \, n \leq 1 \cdot s n^l n \leq 1 \cdot s \cdot 1 \leq s \tag{7}$$

This symbolic reduction can also be expressed graphically, as shown in Eq. 8. In this diagrammatic notation, the elimination of types by means of the inequalities $n \cdot n^r \leq 1$ and $n^l \cdot n \leq 1$ is denoted by a 'cup', while the fact that the type s is retained is represented by a straight wire.

3.2   Categorical Compositional Models

The symbolic account and distributional approaches are linked by the fact that they share the common structure of a compact closed category. This compatibility allows the compositional rules of the grammar to be applied in the vector space model, so that sentences may be mapped into one shared meaning space.

A *compact closed category* is a monoidal category in which for each object A there are left and right dual objects A^l and A^r, and corresponding unit and counit morphisms $\epsilon^l : A^l \otimes A \to I$, $\eta^l : I \to A \otimes A^l$, $\epsilon^r : A \otimes A^r \to I$, $\eta^r : I \to A^r \otimes A$ such that the *snake equations* hold. The morphisms of compact closed categories can be expressed in a convenient graphical calculus [19].

The underlying poset of a pregroup can be viewed as a compact closed category with monoidal structure given by the pregroup monoid, and morphisms $\epsilon^l, \eta^l, \epsilon^r, \eta^r$ witnessing the inequalities of (6). Distributional vector space models live in the compact closed category **FHilb** of finite dimensional real

Hilbert spaces and linear maps. Given a fixed basis $\{\mathbf{v}_i\}_i$ of V, ϵ and η are defined by:

$$\epsilon : V \otimes V \to \mathbb{R} :: \sum_{ij} c_{ij}\mathbf{v}_i \otimes \mathbf{v}_j \mapsto \sum_i c_{ii}, \quad \eta : \mathbb{R} \to V \otimes V :: 1 \mapsto \sum_i \mathbf{v}_i \otimes \mathbf{v}_i$$

3.3 Grammatical Reductions in Vector Spaces

Following [27], reductions of the pregroup grammar may be mapped into the category **FHilb** of finite dimensional Hilbert spaces and linear maps using an appropriate strong monoidal functor $\mathsf{Q} : \mathbf{Preg} \to \mathbf{FHilb}$. Strong monoidal functors automatically preserve the compact closed structure. For $\mathsf{Preg}_{\{n,s\}}$, we must map noun and sentence types to appropriate finite dimensional vector spaces:

$$\mathsf{Q}(n) = N \qquad \mathsf{Q}(s) = S$$

Composite types are then constructed functorially using the corresponding structure in **FHilb**. Each morphism α in the pregroup is mapped to a linear map interpreting sentences of that grammatical type. Then, given word vectors $\mathbf{w}_i$ with types p_i, and a type reduction $\alpha : p_1 p_2 ... p_n \to s$, the meaning of the sentence $w_1 w_2 ... w_n$ is generated by:

$$\mathbf{w}_1 \mathbf{w}_2 ... \mathbf{w}_n = \mathsf{Q}(\alpha)(\mathbf{w}_1 \otimes \mathbf{w}_2 \otimes \cdots \otimes \mathbf{w}_n)$$

For example, as described in Sect. 3.1, transitive verbs have type $n^r s n^l$, and can therefore represented in **FHilb** as a rank-3 space $N \otimes S \otimes N$. The transitive sentence *Clowns tell jokes* has type $n(n^r s n^l)n$, which reduces to the sentence type s via $\epsilon^r \otimes 1_s \otimes \epsilon^l$. So if we represent **tell** by:

$$\mathbf{tell} = \sum_{ijk} c_{ijk}\mathbf{e}_i \otimes \mathbf{s}_j \otimes \mathbf{e}_k$$

using the definitions of the counits in **FHilb** we then have:

$$\mathbf{Clowns\ tell\ jokes} = \epsilon_N \otimes 1_S \otimes \epsilon_N(\mathbf{Clowns} \otimes \mathbf{tell} \otimes \mathbf{jokes})$$
$$= \sum_{ijk} c_{ijk}\langle \mathbf{Clowns}\,|\,\mathbf{e}_i \rangle \otimes \mathbf{s}_j \otimes \langle \mathbf{e}_k\,|\,\mathbf{jokes} \rangle$$
$$= \sum_j \mathbf{s}_j \sum_{ik} c_{ijk}\langle \mathbf{Clowns}\,|\,\mathbf{e}_i \rangle\langle \mathbf{e}_k\,|\,\mathbf{jokes} \rangle$$

This equation has the graphical representation given in 8:

$$(8)$$

Meanings of sentences are compared using the cosine distance between vector representations. Detailed presentations of the ideas in this section are given in [13,27], and an introduction to relevant category theory is provided in [12].

4   Categorical Compositional Cognition

Within both ICS and CATCO, we can view sentence meanings in the following way: the semantics of the individual words of the sentence are given as vectors, and the grammar of the sentence is given as an n-linear map, which is linear in each component. In this section, we map the ICS model to the CATCO model, creating a model for categorical compositional cognition, or CATCOG.

The representation in [32] is of the following form:

$$\sum_i \mathbf{f}_i \otimes \mathbf{r}_i \in \bigoplus_i V \otimes \mathcal{R}_{(i)}$$

The index i here is general, but if we are considering the set of roles to describe a binary tree, then the i corresponds to the depth of the tree.

By using carefully chosen matrices, described in [32], this representation can be written as:

$$\mathbb{W} \cdot \mathbf{f}$$

where here, $\mathbf{f} = \bigoplus \mathbf{f}_i$. This representation allows the sentence to be processed by matrix multiplication, changing order and meaning of words.

In the CATCO model, the representation starts with a tensor product of semantic fillers, represented by triangles in the graphical calculus, and then applies an n-linear map. In order to bring this application in line with the ICS representation, we should represent these fillers as a direct sum. There is a map from a direct sum of vectors to a tensor product of vectors expressed as:

$$\bigoplus_i \mathbf{v}_i \mapsto \bigotimes_i \mathbf{v}_i \tag{9}$$

Given a direct sum of vectors, we firstly convert this to a tensor product of vectors. For example, suppose we have vectors $\mathbf{a} = [a_1, ..., a_n]^\top$ and $\mathbf{b} = [b_1, ..., b_m]^\top$. The direct sum of these is essentially the vector formed by concatenation, $[\mathbf{a}, \mathbf{b}]^\top$. From this, we can map to the tensor product of the two vectors, which in this case can be viewed as the outer product, i.e. the matrix with elements $\{a_i b_j\}_{i=1...n, j=1...m}$. We then apply the n-linear map formed of ϵ, η, 1 that corresponds to the grammatical structure of the sentence. The action of this linear map is the same as tensor contraction, of which matrix-vector multiplication is an instance. This maps the vectors we start out with down to one sentence vector. All such sentence representations inhabit one finite shared meaning space, rather than the unbounded meaning spaces required in [32]. The CATCO model shows how tensor contraction can be used to form sentence meanings in a way that fully utilizes grammatical structure. This comparison is given in diagrammatic form in Fig. 1.

Now that we can see that ICS and CATCO have the same sort of structure, we can cross-fertilize in order to reap the maximum benefit from each representation. The ICS representation has been developed with connectionist implementations in mind, and therefore methods developed in [32], and implementations such as [14–16,33] can be used to develop the CATCO model into a cognitive system rather than the purely linguistic system that it is currently used for.
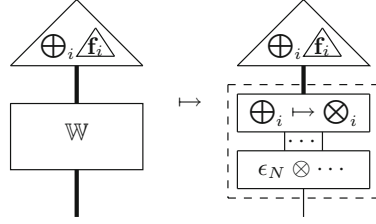
Fig. 1. Comparison of ICS (left) and CatCo (right). In ICS, the string of filler vectors is sent to a linear map which forms a structured tree, represented by thick lines. In CatCo, the string of filler vectors, again represented by a thick line, is first sent to a tensor product of vectors, represented by multiple thin lines. An n-linear map is then applied mapping the string of vectors down to one vector - a single thin line.

ICS is able to characterise a notion of approximate grammatical parsing, called *harmony* that the CatCo model currently lacks. Relationships between the two are explored in [21]. Implementation of the CatCo model within a cognitive framework will also allow for richer representations. Currently, this model is limited to linguistic representations, where meanings of words are derived from text corpora. However, the compositionality that CatCo is able to carry out should transfer very well to meanings derived from other stimuli, such as sound, vision, and so on.

5 Semantic Roles in CatCog

In Fig. 1, we showed how the grammatical structure of a sentence in the CatCo model is viewed as a linear map that corresponds to the matrices used to encode grammatical structure in ICS. However, we may want to enrich the roles from a purely grammatical map to having some semantic content. This would be useful if we wish to utilise the idea of having semantic roles as well as purely formal roles. Semantic roles are discussed in [31,32]. These essentially allow the role vectors to have some semantic content of their own. For example, a proposition 'John gave Mary the book' can be analyzed as having three semantic roles: the giver, the recipient, and the object given. Clearly here, there is more information about the action expressed by the predicate 'give' than is available in purely grammatical roles representing nouns and verbs. This can be implemented within CatCog by manipulating the diagrams we use. We give here the diagram manipulation and a procedure for bringing semantic content into the role.

$$(10)$$

We can also recursively bring more chunks of semantic information into the role vectors if desired. A symbolic structure **s** is represented by a collection of structural roles $\{r_i\}$ represented by $\mathbb{W}_i$ and a base filler $\mathbf{f}_1$ represented by a tensor product of atomic fillers $\{\mathbf{a}_j\}$. The realization of **s** is an activation vector $\mathbf{s} = \mathbb{W}_n \cdot \mathbf{f}_n$ that is the recursive matrix-vector multiplication of a matrix $\mathbb{W}_i$ realizing r_i with a vector $\mathbf{f}_i$ realizing a filler/role binding $\mathbf{f}_{i-1}/r_{i-1}$, i.e.

$$\mathbb{W}_n \cdot \mathbf{f}_n = \mathbb{W}_n \cdot (\mathbb{W}_{n-1} \cdot \mathbf{f}_{n-1}) \tag{11}$$
$$= \mathbb{W}_n \cdot (\cdots \mathbb{W}_1 \cdot (\mathbf{a}_1 \otimes \cdots \otimes \mathbf{a}_m) \cdots) \tag{12}$$

A simple representation, $\mathbb{W} \cdot \mathbf{f}$, where $\mathbf{f}$ is the tensor product of atomic fillers, for a symbolic structure **s** realized by $\mathbb{W}_n \cdot \mathbf{f}_n$ is as follows:

1. Apply an identity matrix (of the appropriate dimension) to each atomic filler in $\mathbf{f}_n$
2. Pull out the matrix-vector multiplication over tensor products to give $\mathbf{f}_n = \mathbb{W}_{n-1} \cdot \mathbf{f}_{n-1}$
3. Repeat steps 1 and 2 recursively until $\mathbf{f}_i$ is the tensor product of atomic fillers $\mathbb{W} \cdot \mathbf{f} = (\mathbb{W}_n \cdot \ldots \cdot \mathbb{W}_1) \cdot (\mathbf{a}_1 \otimes \cdots \otimes \mathbf{a}_m)$

This procedure is equivalent to "stretching up" atomic fillers (e.g. nouns in the linguistic case) and "drawing a box" around them to form $\mathbf{f}$ and then "drawing a box" around the rest of the structure to give $\mathbb{W}$, as shown in (10). For example:

$$\textbf{Clowns tell funny jokes} = \mathbb{W}_{\text{tell}} \cdot (\textbf{Clowns} \otimes [\mathbb{W}_{\text{funny}} \cdot \textbf{jokes}]) \tag{13}$$
$$= \mathbb{W}_{\text{tell}} \cdot (\mathbb{I}_{d_N} \otimes \mathbb{W}_{\text{funny}}) \cdot (\textbf{Clowns} \otimes \textbf{jokes}) \tag{14}$$
$$= \mathbb{W}' \cdot (\textbf{Clowns} \otimes \textbf{jokes}) \tag{15}$$

Using this type of representation, we can also represent relative pronouns such as 'which'. The phrase **Clowns who tell jokes** has string diagram

$$\tag{16}$$

To represent this in the ICS format $\mathbb{W} \cdot \mathbf{f}$, we construct matrices implementing the grammatical structure. In CatCo, the grammatical morphisms are $(\mu_N \otimes \iota_S \otimes \epsilon_N)$ which we apply to the vectors $(\textbf{Clowns} \otimes \textbf{tell} \otimes \textbf{jokes})$.

μ can be thought as a multiplication map that pointwise multiplies two vectors together, and ι can be thought of as a deleting map. For a concrete example, suppose $N = \mathbb{R}^2 = S$ and let $\{\mathbf{n}_i\}_i$ and $\{\mathbf{s}_j\}_j$ denote orthonormal bases of N and S respectively. **Clowns**, **jokes** $\in N$, **tell** $\in N \otimes S \otimes N$ are given in (17). Note that **tell** is a rank-3 tensor with entries $[\textbf{tell}]_{ijk}$. For example, $[\textbf{tell}]_{212} = 9$.

$$\textbf{Clowns} : \begin{bmatrix} 7 \\ 4 \end{bmatrix} \quad \textbf{jokes} : \begin{bmatrix} 5 \\ 1 \end{bmatrix} \quad \textbf{tell} : \begin{bmatrix} 3 & 8 & 4 & 1 \\ 6 & 2 & 9 & 5 \end{bmatrix} \tag{17}$$

From these, matrices are constructed to implement the grammatical structure. $\mathbb{W}_{\mathbf{v}}^{\mu}$ is a matrix that implements the Frobenius multiplication of a vector with $\mathbf{v}$, W_ι is the deleting map, and $\mathbb{W}_{\text{tell}}^{Obj}$ performs application of a verb to an object.

$$\overrightarrow{\textbf{Comedians who tell jokes}} = \mathbb{W}_{\text{Comedians}}^{\mu} \cdot (\mathbb{W}_\iota \cdot (\mathbb{W}_{\text{tell}}^{Obj} \cdot (\textbf{jokes}))) \qquad (18)$$

$$= \begin{bmatrix} 7 & 0 \\ 0 & 4 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 3 & 8 & 4 & 1 \\ 6 & 2 & 9 & 5 \end{bmatrix}^T \cdot \begin{bmatrix} 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 441 \\ 156 \end{bmatrix} \qquad (19)$$

6   Unbinding

The availability of an unbinding mechanism is essential for systematicity in cognitive architectures. This means that given a sentence such as 'Clowns tell jokes', we may access the individual components of the sentence, allowing us to, for example, answer questions about what is going on in the sentence, or allowing us to take the sentence apart and modify it. We propose an approximate unbinding operation which arises naturally from the pulling down of the semantic information into the role information, and hence we use the representation that we introduced in (10).

6.1   Approximate Unbinding

Unbinding is the procedure where we extract a filler from a semantic binding. To achieve this, we require a method to invert $\mathbb{W}_r$, which may not be invertible. We therefore use the Moore-Penrose pseudoinverse for approximate unbinding. For a binding $\mathbb{W}_r \cdot \mathbf{f}$, we approximately unbind $\mathbf{f}$ from $\mathbb{W}_r$ by application of the Moore-Penrose pseudo inverse of $\mathbb{W}_r$: $\mathbb{W}_r^+ \cdot (\mathbb{W}_r \cdot \mathbf{f}) \approx \mathbf{f}$.

Consider $\mathbf{s} = \mathbb{W}_{\text{tell}} \cdot (\textbf{Clowns} \otimes \textbf{jokes})$ where $\textbf{Clowns}, \textbf{jokes} \in N$. We want to insert the adjective *funny*, giving $\mathbf{t}$:

$$\mathbf{t} = \mathbb{W}_{\text{tell}} \cdot (\textbf{Clowns} \otimes [\mathbb{W}_{\text{funny}} \cdot \textbf{jokes}]) \qquad (20)$$

This mapping can be done using $\mathbb{W}_{\mathcal{F}} \cdot \mathbf{s}$ where $\mathbb{W}_{\mathcal{F}} = \mathbb{W}_{\text{tell}} \cdot (\mathbb{I}_{d_N} \otimes \mathbb{W}_{\text{funny}}) \cdot \mathbb{W}_{\text{tell}}^+$

$$\mathbb{W}_{\mathcal{F}} \cdot \mathbf{s} = [\mathbb{W}_{\text{tell}} \cdot (\mathbb{I}_{d_N} \otimes \mathbb{W}_{\text{funny}}) \cdot \mathbb{W}_{\text{tell}}^+] \cdot \mathbf{s} \qquad (21)$$

$$\approx \mathbb{W}_{\text{tell}} \cdot (\mathbb{I}_{d_N} \otimes \mathbb{W}_{\text{funny}}) \cdot \mathbb{I}_{d_N \cdot d_N} \cdot (\textbf{Clowns} \otimes \textbf{jokes}) \qquad (22)$$

$$= \mathbb{W}_{\text{tell}} \cdot (\mathbb{I}_{d_N} \otimes \mathbb{W}_{\text{funny}}) \cdot (\textbf{Clowns} \otimes \textbf{jokes}) \qquad (23)$$

$$= \mathbb{W}_{\text{tell}} \cdot (\textbf{Clowns} \otimes [\mathbb{W}_{\text{funny}} \cdot \textbf{jokes}]) = \mathbf{t} \qquad (24)$$

7   Consequences of CatCog

The structure of the CatCog model means that all the power of categorical compositional semantics can be leveraged to represent phenomena that are useful in a model of cognitive AI. We give a short description of these structures and how they will be useful.

The representation of ambiguity is important in cognition. How can one representation mean more than one thing, and how does context collapse the ambiguous symbol down to one meaning? CatCo uses quantum theory which has a ready-made structure called a density matrix that can represent ambiguous symbols. These can be used with grammatical composition to model word ambiguity and how the sentence context disambiguates [17,24,25].

Density matrices have also been used to implement logical entailment at the word and the sentence level [1–4]. Results in these papers show how logical entailment between two sentences can be derived as a function of logical entailments between the words in the sentences, within a distributional representation. These results are useful for implementation of logical reasoning, and showing how reasoning at the sentence level will work.

More subtle grammatical structures can be represented such as relative pronouns. Using relative pronouns allows us to form definitional noun phrases such as 'The woman who rules England'. These noun phrases are represented in the same noun space N as their components, and we can therefore compare them directly [18,28,29]. This will be useful in modelling knowledge update.

The CatCo representation has also been used in examining how psychological phenomena such as over- and under-extension of concepts can occur [3,11]. This is a particularly interesting area, since it is not clear that these type of phenomena can be adequately represented using ICS style representations. We give an example to illustrate.

Suppose we take the vectors **pet** and **fish** and suppose we choose role vectors $\mathbf{mod} = [1,0]^\top$ and $\mathbf{noun} = [0,1]^\top$. In the ICS representation, we have

$$\mathbf{pet\ fish} = \mathbf{pet} \otimes \mathbf{mod} + \mathbf{fish} \otimes \mathbf{noun}$$

Then, we may wish to compare **goldfish** and **pet fish**. We cannot do this directly without some initial processing. One option might be to form the tensor product **goldfish**⊗**noun**, and use the matrix inner product. However, then, the similarity depends only on the similarity between the noun **goldfish** and the noun **fish**, due to the orthogonality of the role vectors. An objection to this might be that the role vectors do not have to be completely orthogonal, but may be noisy. Then, the similarity is just a noisy similarity to **fish**, and **pet** still plays no real role in the combined meaning.

Another approach would be to use circular convolution [26]. Applying circular convolution to each of **pet**⊗**mod** and **fish**⊗**noun** leaves **pet** as is, but shifts the indices of **fish** by one. These vectors are then summed. If we view the dimensions of the vectors **pet** and **fish** as representing attributes, this gives a notion of interaction between attributes. However, this notion seems ad hoc - why should the indices of the attributes be shifted along by one? In contrast, the CatCo model uses the underlying grammatical structure of the sentence to explain the interactions of attributes [11].

The abstract framework of the categorical compositional scheme is actually broader in scope than natural language applications. It can be applied in other settings in which we wish to compose meanings in a principled manner, guided by

structure. Another extension is therefore in using other representational formats such as conceptual spaces [5].

8   Conclusion and Outlook

So what have we achieved with our CatCog representation in relation to ICS? The benefit of this new recursive representation is that filler/role bindings (i.e. constituents) that make up a symbolic structure **s** are now composed in such a way that all well-formed **s**, with respect to a certain cognitive task, are realized in a *finite shared* meaning space. This allows the comparison of well-formed symbolic structures with different underlying grammatical structures.

This new representation opens up a number of avenues for further work. On the theoretical side, a key line of enquiry will be to push the comparison between ICS and CatCo further. This will allow us to analyse the type of theoretical structure that is used within the $\mathbb{W}$ matrices employed by ICS. In particular, the representations of verbs, adjectives, and other relational words in CatCo inhabit a higher dimensional space than nouns, and therefore it might be thought that there is an unfair comparison between the two models. In fact, it is possible to take a vector representation, and lift it into a higher dimensional space. Investigating how to do this so ICS structure is preserved is future work.

Another current line of research will look at how to properly formalise the notion of knowledge updating. If I tell you that *John runs*, and you previously did not know this, how is your representation of *John* updated? Again, architectures including ICS and [16] will feed into this research.

On the implementation side, future work in this area will be to apply the theory within a model such as Nengo [14] or LISA [16]. These implementations already use tensor product representations, and therefore have the right kind of underlying structure to serve as a good implementation. Extensions of approaches such as [22,31] will also be fruitful.

CatCog draws inspiration from categorical quantum mechanics, and therefore techniques and structures from quantum theory can be incorporated into the formalism. Further uses concern phases and (strong) complementarity [9].

While our examples are all linguistic, our model accounts for general cognitive tasks that manipulate filler and roles. We therefore leave with a programme for producing compositional structure within distributed representation of cognitive processes.

References

1. Balkır, E.: Using density matrices in a compositional distributional model of meaning. Master's thesis, University of Oxford (2014)
2. Balkır, E., Kartsaklis, D., Sadrzadeh, M.: Sentence entailment in compositional distributional semantics. In: ISAIM 2016 (2015, to appear)
3. Bankova, D.: Comparing meaning in language and cognition. Master's thesis, University of Oxford (2015)

4. Bankova, D., Coecke, B., Lewis, M., Marsden, D.: Graded entailment for compositional distributional semantics (2016). arXiv:1601.04908
5. Bolt, J., Coecke, B., Genovese, F., Lewis, M., Marsden, D., Piedeleu, R.: Interacting conceptual spaces (2016, Submitted)
6. Clark, S., Coecke, B., Grefenstette, E., Pulman, S., Sadrzadeh, M.: A quantum teleportation inspired algorithm produces sentence meaning from word meaning and grammatical structure. Malays. J. Math. Sci. **8**, 15–25 (2014). arXiv:1305.0556
7. Clark, S., Pulman, S.: Combining symbolic and distributional models of meaning. In: Quantum Interaction, pp. 52–55 (2007)
8. Coecke, B.: The logic of quantum mechanics – take II. In: Chubb, J., Eskandarian, A., Harizanov, V. (eds.) Logic and Algebraic Structures in Quantum Computing, pp. 174–198. Cambridge University Press (2016). arXiv:1204.3458
9. Coecke, B., Duncan, R.: Interacting quantum observables: categorical algebra and diagrammatics. New J. Phys. **13**, 043016 (2011). arXiv:quantph/09064725
10. Coecke, B., Grefenstette, E., Sadrzadeh, M.: Lambek vs Lambek: functorial vector space semantics and string diagrams for Lambek calculus. Ann. Pure Appl. Logic **164**, 1079–1100 (2013)
11. Coecke, B., Lewis, M.: A compositional explanation of the 'Pet Fish' phenomenon. In: Atmanspacher, H., Filk, T., Pothos, E. (eds.) QI 2015. LNCS, vol. 9535, pp. 179–192. Springer, Heidelberg (2016). doi:10.1007/978-3-319-28675-4_14
12. Coecke, B., Paquette, E.: Categories for the practising physicist (2009). arXiv:0905.3010
13. Coecke, B., Sadrzadeh, M., Clark, S.: Mathematical foundations for a compositional distributional model of meaning. Linguist. Anal. **36**(1–4), 345–384 (2011)
14. Eliasmith, C.: How to Build a Brain: A Neural Architecture for Biological Cognition. OUP, Oxford (2013)
15. Eliasmith, C., Thagard, P.: Integrating structure and meaning: a distributed model of analogical mapping. Cognit. Sci. **25**(2), 245–286 (2001)
16. Hummel, J.E., Holyoak, K.J.: Distributed representations of structure: a theory of analogical access and mapping. Psychol. Rev. **104**(3), 427 (1997)
17. Kartsaklis, D.: Compositional distributional semantics with compact closed categories and frobenius algebras. Ph.D. thesis, University of Oxford (2015)
18. Kartsaklis, D., Sadrzadeh, M., Pulman, S., Coecke, B.: Reasoning about meaning in natural language with compact closed categories and Frobenius algebras. In: Logic and Algebraic Structures in Quantum Computing, p. 199 (2013)
19. Kelly, M., Laplaza, M.L.: Coherence for compact closed categories. J. Pure Appl. Algebra **19**, 193–213 (1980)
20. Lambek, J.: Type grammar revisited. Log. Asp. Comput. Linguist. **1582**, 1–27 (1999)
21. Lewis, M., Coecke, B.: Harmonic grammar in the disco model of meaning, ADS@IWCS (2015)
22. Milajevs, D., Kartsaklis, D., Sadrzadeh, M., Purver, M.: Evaluating neural word representations in tensor-based compositional settings. EMNLP (2014)
23. Montague, R.: The proper treatment of quantification in ordinary english. In: Hintikka, K.J.J., Moravcsik, J.M.E., Suppes, P. (eds.) Approaches to natural language, pp. 221–242. Springer, Heidelberg (1973)
24. Piedeleu, R.: Ambiguity in categorical models of meaning. Master's thesis, University of Oxford (2014)
25. Piedeleu, R., Kartsaklis, D., Coecke, B., Sadrzadeh, M.: Open system categorical quantum semantics in natural language processing. CALCO (2015)

26. Plate, T.: Holographic reduced representations. Neural Netw. **6**(3), 623–641 (1995)
27. Preller, A., Sadrzadeh, M.: Bell states and negative sentences in the distributed model of meaning. ENTCS **270**(2), 141–153 (2011)
28. Sadrzadeh, M., Clark, S., Coecke, B.: The Frobenius anatomy of word meanings I. J. Logic Comput., ext044 (2013)
29. Sadrzadeh, M., Clark, S., Coecke, B.: The Frobenius anatomy of word meanings II. J. Logic Comput., exu027 (2014)
30. Schütze, H.: Automatic word sense discrimination. Comput. Linguist. **24**(1), 97–123 (1998)
31. Shastri, L., Ajjanagadde, V.: From simple associations to systematic reasoning. Behav. Brain Sci. **16**(03), 417–451 (1993)
32. Smolensky, P., Legendre, G.: The Harmonic Mind. MIT Press, Cambridge (2005)
33. Thagard, P., Stewart, T.C.: The AHA! experience: creativity through emergent binding in neural networks. Cognit. Sci. **35**(1), 1–33 (2011)
34. Turney, P.D., Pantel, P., et al.: From frequency to meaning: vector space models of semantics. J. Artif. Intell. Res. **37**(1), 141–188 (2010)