Modeling Meaning: A Kantian Intervention in Vector Space Semantics

by

Nipun Arora

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

Master of Cognitive Science

in

Cognitive Science

Carleton University

Ottawa, Ontario

Abstract

This thesis discusses the implementation of a set of logical forms to enrich the way meaning is modeled in a vector-based system of conceptual memory. Vector-space models can account for a variety of psycho-linguistic phenomena by representing relationships between concepts as distance in a high-dimensional space. But they lack logical organizational structure without which inferential operations are impossible. Augmenting cognitive architectures with innate, logical structures might be the key to resolving this issue. But proposing such structures risks over-attributing the complexity of behavior to complexity in the architecture. I propose using Kant's critical work for a strong theory to select a minimal set of logical forms. The Kantian logical forms are implemented onto vector space architecture in a system (Kantian-HDM) created in R programming language and has been published on GitHub. The results of the simulations run in the system are presented along with a description of the inferential behavior exhibited.

Acknowledgment

First and foremost, I would like to thank my co-supervisors Andrew Brook and Robert West. Without Andrew's brilliant course on Kant, I would never have been able to give this work the philosophical direction vital to it. The many discussions I had with Rob were crucially important to stay motivated as I navigated the highly intersectional domain of this work. Their support in preparation and writing of this thesis was critical for its success. Second, I am grateful to Raj Singh for introducing me to the domain of computational linguistics and providing key feedback for this work. Thirdly, I want to extend my gratitude to Mathew Kelly who made himself available to help me through the often obscure literature and models on the topic of this thesis.

I must also take a moment to thank my parents and sister who never faltered in their support for my pursuits. Finally, I would like to thank Garima Arora, whose consistent presence and encouragement was crucial to maintain my spirits for the duration of this work.

Contents

List of Tables

List of Illustrations

Introduction

Meaningfulness arises as one of the most fundamental of experiences in our interaction with the world and underlies all our inquiries—both philosophical and mundane. What often occurs in the attempts to study such a complex phenomenon is a deconstructive process aimed to understand it through the study of more manageable sub-pieces. The literature on meaning thus identifies many heterogeneous features of this seemingly homogenous experience. Consequently, investigations on meaning is spread across a diverse range of fields of research including philosophy, psychology, language and more recently cognition. The different findings and theoretical emphases of these disciplines have led to a body of constraints which while enormously difficult to satisfy at the same time, are also necessary for any model of meaning which aims to do the phenomenon its due justice.

Philosophers studying meaning have traditionally focused on the question of conceptual semantics and possession conditions of concepts (and knowledge in general). This has led to theorization over how we acquire meaning, research on perception as means of acquiring knowledge, and investigations into truth-conditions of the acquired knowledge. The philosophers tackle directly the question of how meaning and knowledge interface with sensory systems which collect the information about the world. While some philosophers have tried to formulate the investigations outward in the study of the world which meaning refers to, others have directed their queries inwards towards the mind which creates this meaning. As a result, we can roughly divide philosophy's dealing with meaning in two—studying the conditions which allow us to have mental representation/symbols, and the world which grounds these.

Linguistics, on the other hand, has focused on *formal-logical* features of meaningful linguistic expressions, and by extension, mental representations which these expressions stand for, sometimes borrowing from classical logic. The former consists of rules and relationships which are required to capture the aspects of meaning that can be recognized as symbol transformations and information processing. These include set-theoretic relations, compositionality, recursive binding, etc. which have been realized in some semantics models through formal structures like function and argument systems, quantifiers and scope, etc.

Psychology has been usually interested in the behavior that arises around meaning and the *psycho-linguistic* constraints. These arise due to the fact that meaning is not a mind-independent phenomenon, but rather is a cognitive experience deployed by humans during their interactions with the world. The focus thus is on tasks like categorization and recall. To model this aspect of meaning requires modeling constraints of both storage and behavior of stored knowledge in the mind. These include the emergence of the psychological phenomena around meaning such as the gradience in the similarity of concepts, recall and identification errors, synonymy tests, effects of semantic priming, etc. These are thus concerned with the sub-symbolic aspects of meaning. A harder version of such a constraint is imposed by neuroscience which arises due to the fact that meaning is a neural phenomenon. Neuroscience imposes a general constraint over all cognitive models that cognitive models must deploy operations and symbol transformations which can be realized in a neural architecture. Otherwise they cannot be understood as proper models of meaning in humans. The commitment this asks of the models is that to not merely replicate the results but the manner.

Often the computational/symbolic aspects of meaning are categorized under *competence system* and the sub-symbolic ones are considered part of the *performance system*, separate from the competence system. Such a competence-performance distinction is important for linguists as it allows them to protect theoretical models of language from issues such as speech errors which occur when those models get executed to produce linguistic behavior. Psychological phenomenon such as errors are considered to arise due to mitigating circumstances which play a role in the production of language behavior (performance), and thus do not require to be incorporated in the core linguistic competency. The roots of such a divide can be found in Kant's critical work (Kant, 1781) where he divided the study of thought into *general logic* and *applied logic*. For Kant, general logic concerns the rules that are absolutely necessary for understanding. In it, "we abstract from all empirical conditions under which we exercise our understanding…from influence of the senses, from the play of imagination, from the laws of memory, from the force of habit, from inclination, etc."[1] Applied logic, on the other hand, is the study of general logic "put to use in subjective empirical conditions taught us by psychology". Hence, things such as attention, memory, desires, goals, emotions, etc., come into the picture.

[1] Critique of Pure Reason, A 53/B 77

Because the differences in the nature of these constraints are perceived to be fundamental, most of the models of meaning address only one kind of constraints. The logic-based semantic models like lambda-calculus are designed to implement the formal-logical constraints which allow for systematic integration of concepts. On the other hand, the *distributional-statistical* models are good at modeling the psycho-logical constraints and have architecture options which are neurally plausible.

Since Newell (1973), there has been a call for the unification of the siloed theories of cognition. This comes from the concern Newell had for the future of psychology. Psychological research for him seemed to be generating detailed clarification of many psychological phenomena through an approach he described as "playing the game of *twenty questions* with nature". This was troublesome for Newell as it resulted in a vast amount of knowledge that was difficult to put together. Often divisions like those of competence system and performance system "harden…into a firewall: competence theories [have come] to be considered immune to evidence from performance" (Jackendoff, 2007b). In order to develop a unified theory of meaning, it is important then that divisions like these are kept methodological, and the research occurs dialectically between the different fields rather than independently. For Sellars (1962), it is "the aim of philosophy, abstractly formulated… to understand how things, in the broadest possible sense of the term, hang together in the broadest possible sense of the term". This thesis discusses an attempt at exactly this kind of unification by implementing logical forms in vector-based distributional models which have been previously shown to model psycho-linguistics behavior well. Some models attempt to use vector-based representations as input for a separate

logical-semantic construction. This thesis, inspired by Kantian philosophy, concerns itself with use of a certain kind of logical form as unifying functions through which representations are created in the first place. Implementation of such domain-general ontology within the vector-model themselves has not been done before.

The vector-based models of meaning (for a detailed review refer to Jeff Mitchell & Lapata, 2008), which fall within distributional-statistical semantics, have gained popularity due to their neurally plausible architecture, and ability to model psycho-linguistics phenomena including semantic priming, similarity, and association. In these models, each concept is represented by and stored in the form of vectors—sets of numbers like {1,4,2,5….}. The relationships between the concepts are stored by moving their vectors through a high-dimensional space in which they exist. While both compositional and recursive, these models have a weakness—a lack of logical structure which provides a uniform framework to integrate concepts. The same is addressed in this work. The thesis discusses a system coded in a programming language called R created by me to infuse the concepts in vector-space with logical forms. The forms are inspired in their use to create new concepts from the Kantian philosophy which understands them as functions of judgments through which inputs of perception are referred to, and through reference, understood/assimilated in terms of our knowledge. An obstacle which makes such an attempt especially difficult is lack of mapping between the mathematical language of the vector-space architectures and the commonly understood cognitive phenomenon. As a result, there exist many tricks in the literature without a proper methodology of how different models deploying them should be investigated and compared. I address this issue first and then proceed to develop

a system which implements models of vector space which organize themselves based on the Kantian categories.

The term *semantics* is used differently by different disciplines. These differences come from their specific goals or methodologies which guides their theorization. Given the interdisciplinary nature of this thesis, *semantics* is defined as an overarching study for all the necessary constraints, and the methods of their satisfaction, which are necessary to model the experience of meaning in its entirety. The final product of the research work presented here is a system which specifically implements the concept-formation aspect of semantics, but does so in a way that it is possible to bridge it with other domains of semantics like perception. The thesis is organized into seven chapters.

- Chapter 1: The chapter discusses different aspects of meaning that any model of meaning must keep in mind. The chapter is primarily a stage-setting exercise to clarify the goals and method of this work.
- Chapter 2: The chapter develops a body of constraints— computational, psychological and perceptual—that any system aiming to model meaning in its entirety must satisfy.
- Chapter 3: The chapter discusses various models of semantics in the literature and how they fare in regard to the constraints discussed in the previous chapter.
- Chapter 4: The chapter is motivated by the lack of a general-purpose architecture that can satisfy all of the above-mentioned constraints. It identifies Vector Space Architecture (VSA) as a promising solution to fill this lack. The chapter introduces VSA and a specific model developed in the Cognitive Modelling Lab at Carleton. It then identifies as weaknesses the lack of clarity around the mathematical properties of VSAs, the lack of methodological evaluation of these model of VSA, and the lack of logical forms in these models.

- Chapter 5: In this chapter, acknowledging the necessary requirement of a better understanding of VSAs in order to develop a method to deploy logical forms in it, I discuss the simulations I conducted to shed more light on the specific mechanics of the architecture.
- Chapter 6: In this chapter, I discuss the Kantian epistemology and make a case for why Kantian categories make the best fit for implementation of logical forms in VSA models. I provide the assessment of Kantian theory to motivate its use for the model.
- Chapter 7: The chapter discusses a new system (Kantian-HDM) which I created in a programming language called R. It implements Kantian Categories in vector-space architecture and allows for a systematized way of integrating concepts to form new ones, thus laying the ground for general inferential abilities in such models. I discuss the kind of inference abilities which Kantian-HDM exhibits by examining the results of the simulations I ran l. I finally conclude with future directions for the work.

Chapter 1: Meaning's Meaning

1.0 Chapter Overview

The chapter discusses different aspects of meaning—form, content, expression that any models of meaning must keep in mind. The chapter is primarily a stage-setting exercise to clarify the kind of semantics I have in mind, and identifies the goals and method of this work. The chapter also delineates how different disciplines invested in the study of meaning relate to one another.

1.1 The Domain of Meaning

A theory of meaning must answer two different questions which are often conflated—what the meaning of something is, and what the conditions for something to have meaning are. The purpose of the chapter to delineate and clarify these two aspects of meaning and understand the extent to which the historical and the contemporary treatment of meaning addresses them.

Responses to the first question—what the meaning of something is—are often built around semantic theories of language expressions. Meaning is defined by what the words and expressions correspond to. This correspondence when defined with respect to the description contained in the expression gives us the *sense* of the expression. On the other hand, when it is defined with respect to the external objects, it gives us the *reference* of the expression. The sense of the expression allows us to locate this reference. As a result, expressions can have different senses but the same reference. For example, while the sentences "*my mother"* and "*my father's wife"* may both refer to the same person, each of them would do so via two different senses. Another way in which linguists look at meaning is through the lens of truth conditions and

possible worlds. In this paradigm, the meaning of an expression can be its *Intension*—the set of all things in all possible worlds for which the expression is true; and its *Extension* is set of all things in this world for which it is true. As a result, meaning theory is given a instrumentalist treatment used merely for "systematizing the behavior of linguistic creatures" (Antony & Davies, 1997). For psychological theories of language, sense or intensional meaning is stored in mental representations of objects. Research on embodied meaning insists that the meaning of expressions comes from a sensorimotor experience which is either caused by our interaction with the external objects which the extension refers to, or a simulation of such an interaction. If this is true, words have meaning through mental conceptual representations, which, in turn, get meaning from sensorimotor simulations of the world conditions they refer to.

But the story can be told the other way around as well. Meaningfulness also accompanies perceptions of the world. This raises the question of where the meaning of the perceptual experiences comes from. One must look at the work done in perception studies to better understand this. Treisman (1998) proposes that when sensing an object, we pick out features which are later combined to create the perception of the object. Taking this further, in their paper titled *Understanding What We See: How We Derive Meaning*, Clarke & Tyler (2015) make a case that "Recognizing objects goes beyond vision and requires models that incorporate different aspects of meaning...from low-level visual input through the categorical organization to specific conceptual representations". Their paper makes a case for a feature-based theory in which concepts are composed out of features—shared features lead to the categorization of objects under superordinate concepts (e.g. *animal*) and distinct features lead to basic level concepts (e.g.

dog). At the same time, the formed concepts play a significant role in giving perceptions a highly contextual meaning such that depending on one's currents goals and personal sets of concepts, the same object can be perceived differently. One might look at a piece of paper and see a potential letter or potential fuel for fire depending on what one's current needs or previous experience with paper are. Thus, perceptions also seem to get meaning from conceptual mental representations. This entails, together with the earlier discussion, that meaningfulness can be best understood as a translation of both linguistic and non-linguistic inputs into conceptual thought which mediates bi-directional movement from perception to words (Figure 1). Another way to understand this is as the rendering of perceptions (linguistic and non-linguistic) into conceptual objects.

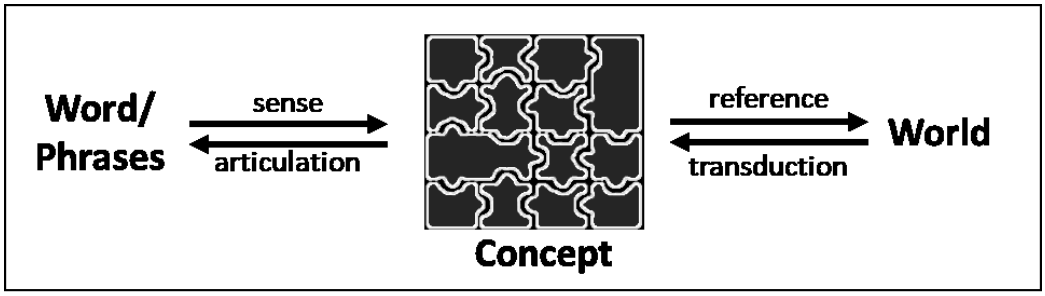

Figure 1: Meaning (Content, Expression)

But a theory of meaning as *reference* of words to the world, through a *sense* constituted by the sensory-motor representations and conceptual representations does not tell us what the necessary conditions to have meaning are. There is what has been referred to as the *meaning barrier* (Chalmers, French, & Hofstadter, 1992) between meaning conceptualized as high-level representations like linguistic expression which allow for discursive knowledge; and low-level perceptions from which these representations are constructed. "The task for psychological

semantics is neither to relate language to a model nor to relate it directly to the world. It is rather to show how language and the world are related to one another in the human mind, to show how the mental representation of sentences is related to the mental representation of the world" (Johnson-Laird, 1982). To locate meaning in merely the empirical experience from which the representations get their content conflates the *explanation of meaning* (how meaning arises) with the *description of the object* to which it refers (i.e., what the contents of meaning are).

The answer to the explanation of meaning requires an additional investigation into the conditions that are necessary to have meaning. One of the primary concern of this thesis is regarding the internal structure of mind which provides forms through which the representations of the data of the external world gets constructed so as to form knowledge of the world which in turn gives rise to the meaningfulness of our experience of the world. I discuss this constructive aspect of meaning in more details in next section.

1.2 Construction of Meaning

The discourse on meaning is spread across aspects of this framework (Figure 1) and is lacking in how it comes together. The linguistic approach to meaning is concerned with how words and expressions signify the contents of their meaning, i.e., entities that they denote. The predominance of a linguistic approach to meaning is only a 20^{th}-century phenomenon. The philosophy of meaning before was securely tied to the acquisition of knowledge through perceptual experience of the world. Recent research on embodied meaning marks a reversion to this flesh and bone approach to meaning. The claim they make is that concepts are stored in the form of specific sensory-motor symbols and it is this particular embodied nature of the symbol

which gives concepts their meaning. Both these approaches are empiricist in that they locate the source of the data that constitutes knowledge in the external world.

But for Philosopher Immanuel Kant, the empiricists wrongly assume that the order among the representations of an experiential manifold[2] is self-evident in our first experience and can be extracted from it. Instead, he claims that representations are properly understood, not as abstracted copies of impressions, but rather as the products of a constructive process undertaken through *schematized categorial rules* of determination which create an objective experience. "These general rules function as the true original constitutive conditions of all experience by arranging, organizing, and integrating the present particulars into a lawfully ordered whole. Accordingly, such a constitution is not an ontological productivity that claims to bring forth the existence of the thing, but a formal constitution that produces the order through which we can know the object as standing within a system, as coherent epitome of laws of nature." (Sloboda, 1995). For Kant, this innate structure constitutes the grounding conditions from which the phenomenon of having knowledge of the world and hence the conditions to possess meaning arises. These functions construct meaning by giving knowledge its significative ability by relating concepts to either other concepts or to non-conceptual (sensory) cognitions (Hanna, 2017).

[2] The diverse collection of data which we are exposed to when we perceive the world

Uncovering these conditions of meaning which allow for its construction entails theorizing about the ontology of the structure of human mind itself. This issue of construction of meaning has cropped up frequently in many areas of research with some variations in how it is defined, and goes by many names—the problem of unity (Kant, 1781), the binding problem (Jackendoff, 2002; Treisman, 1998), semantic unification (Hagoort, Baggio, & Willems, 2009), etc., and has led to postulation of areas in the brain responsible for it. These areas have also been given many names like productive imagination (Kant, 1781), faculty of language Narrow (Hauser, Chomsky, & Fitch, 2002), semantic hub (Patterson, Nestor, & Rogers, 2007), convergence zones (Damasio, Tranel, Grabowski, Adolphs, & Damasio, 2004), etc.

Kant approached this issue by asking what is left in our cognition once we take away the content sourced from experience. Whatever is the leftover from this exercise cannot be something that the world, through our experience of it, contributed; but rather would be what our mind brings to the meaning. He came up with forms—*pure form of understanding*—through which we structure our perceptions and relate concepts. Take this expression— '*A* is *B* but might not be *C*. All *D* can cause *A*. Some *D* must not be *E* …' and so on. This expression does not give us any content or anything we can use to pick out an object in the world, yet it does not sound like gibberish. This is not the same as "*A* kicks *B*…" in which case *A* and *B* are not variables we can say nothing about. We rely on the perceptual experience of *kicking* and objects which *kick*, and so can still rule out objects in the world that cannot kick or be kicked. Similarly, for Kant, no perceptions can be imagined without some spatio-temporal forms—*pure forms of intuition*.

These forms (of intuition and understanding) then constitute an internal structure to meaning for Kant.

We thus have an idea of meaning to go forward that draws from three things—internal structure providing forms, the perceptual content providing the (content) for the concepts, and the linguistic environment, which provides a way for the meaning to be expressed. But, all of these need to a general architecture of the mind and cognition over which these are implemented. This forms the metaphysical non-semantic facts upon which meaning facts supervene. The entire picture is presented in **Figure 2**. The work in this thesis proposes Vector Space Architecture as the psychological architecture, and Kantian Categories as the conceptual forms in this model of semantics.

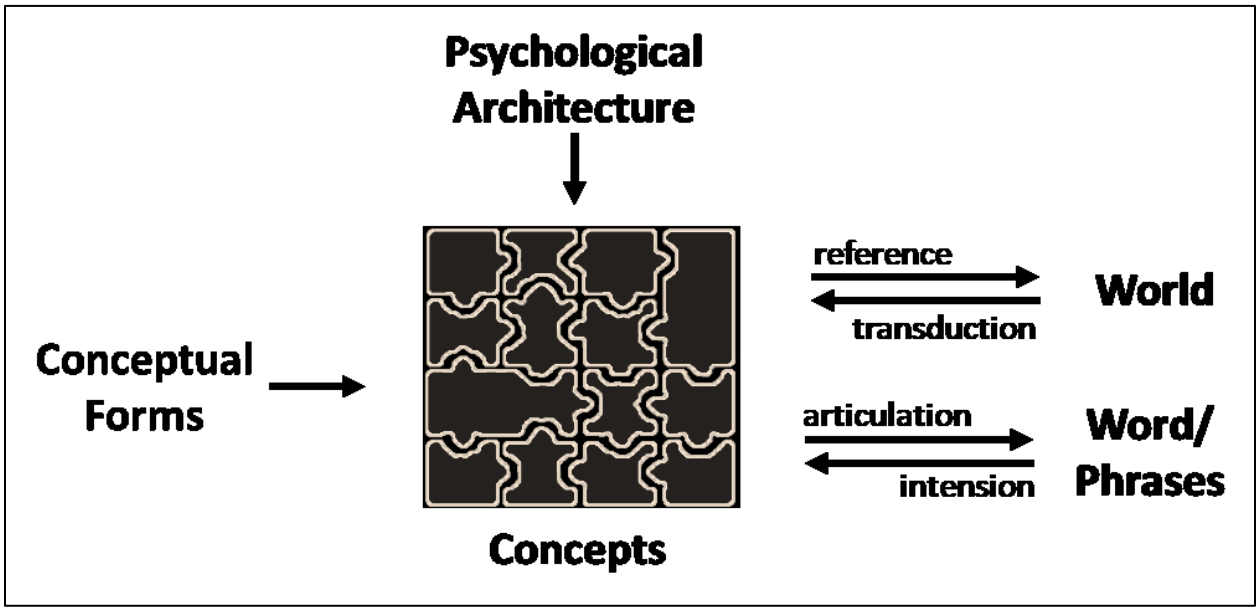

Figure 2: Meaning (Form, Content, Expression)

1.3 Chapter Conclusion

To model meaningfulness in this picture would require, in addition to a theory of what the meaning of linguistic expressions is and the world from which they get their contents, a theory of the forms and dynamics of conceptual mental representations in which these contents get organized and structured. Nevertheless, such a theory needs to be constructed keeping in mind the interfaces it must have with the external world—the information from which it organizes—and language through which it gets expressed. “Any idealization of speakers' knowledge had better be principled, governed not by a priori surmise about the nature of linguistic behavior, but rather by an appreciation of the epistemically relevant features of the learning situation. In short, we need to do psychology in order to do semantics…”(Antony & Davies, 1997). This thesis concerns itself with exactly this task—to locate first the nature of the forms and dynamics of the mind’s knowledge-creating ability, identify the modeling paradigm most suitable to implement these in and make first steps towards such an implementation.

Chapter 2: Constraining the Models of Meaning

2.0 Chapter Overview

The chapter undertakes the development of a body of constraints that must be satisfied by models of meaning in humans. It first discusses the constraint-based approach including the different kinds of constraints. It then proceeds to develop specific constraints which a theory of meaning must satisfy. A significant focus of the chapter is on distinguishing the *psychology through language* from the *psychology of language*, and delineating what each of them can contribute to the theory of meaning.

2.1 On the Importance of Constraints

There is a certain amount of work which models of meaning must be able to do in order be of significance. One of the most significant methodological innovation has been the *transcendental method*—a question of what are the necessary conditions (constraints) which a system must satisfy in order to produce the observed behavior (Brook, 1994). A somewhat weaker version of this approach is also known as *inference to the best explanation* wherein one asks the question of what are the most probable conditions that would have produced the observed behavior. While widely accepted in the practices of research in cognitive science, the method requires us to be acutely mindful of what it is that we are attributing as the behavior of the system we are trying to explain. Thus, it is important to identify the significant markers of the nature of the experience of meaning in humans, and then use these markers to constrain any system which aims to model meaning.

There arises a distinction between these constraints which is similar to the linguistics competence-performance distinction and is especially prevalent in cognitive modeling. This distinction is between the features that play a role in designing the model, and those which are used to test/select the model. At the computational level of analysis, formal methods are used to prove what kinds of problems can and cannot be solved by a particular model. And thus, the phenomenon (say of meaningfulness) is first translated into information processing terms and then modeled. While these abilities constitute what the models must necessarily do, and are necessary to design the model, they don't sufficiently constraint the models. Because these constraints are based primarily on inputs to output functions, we face the issue that in principle, an infinite number of different algorithms exist for computing a single input-output mapping of interest (Johnson-Laird, 1983). Here, *artifacts* like time taken to reach an output, the kinds of errors generated by model, etc., become a way of validating that the model not only gives the required output, but does so in a manner similar to the natural phenomenon being modeled. For example, while there are many models which successfully sort and order a list, each of them, courtesy of the difference in the algorithms they deploy, do so in different amount of times. If one is to model how human mind sorts and order lists, it becomes important not to merely select the model which is the fastest, or the most optimized one, but the one which does it the way the humans do it. This can be tested by looking at the fit between the computational models and human subjects with respect to the *artifacts* like time taken to sort lists of different lengths. This distinction of *design constraints* and *artifactual constraints* is important to keep in mind while selecting which models accurately simulate human behavior and cognition. The distinction is similar to the negative doctrine of functionalism—that function does not determine form. While

Kant has been recognized as a functionalist, his commitment to this negative doctrine has often been neglected (Brook, 1994).

Keeping this in mind, one can begin to look at different approaches to studying meaning such as formal semantics, perception, psycholinguistic behavioral studies etc., which have generated a varied list of features which need to be accommodated in models of meaning. Some of these are design constraints and thus define what any model of meaning must be able to do at the minimum, while others are artifacts which allow us to select the model closest to how humans experience meaning.

2.2 Language and Meaning

> "How astonishing it is that language can almost mean, and frightening that it does not quite."
>
> —Jack Hilbert

Language has consistently been proposed as the uniquely human cognitive ability which sets us apart from the other animals. Many contending theories exist around what must be classified as a language. But whatever these theories may be, arguably, language is interesting, first and foremost, in virtue of its ability to carry meaning, and because of that, be able to be used to communicate it (Hagoort et al., 2009). For Pinker, it is exactly this use of communication for which language evolved (Pinker, 1994). Since language is the most explicit way in which we exhibit conceptual knowledge, most attempts to model this knowledge are based on language artifacts, and thus use amodal, symbolic representations. Another reason to prefer language-like representations is that linguistics as an enterprise has been particularly fruitful in formalizing

some of the key features of human cognition including compositionality and generativity. Choosing to study and model meaning through meaningful linguistic expressions is a significant choice and not without consequences. Given that language is the tool for discourse, the conflation of the media with the content is nowhere more prevalent than it is in studying the structure of thought through an analysis of the structure of language which is used to express that thought. What this means is that it is easy to conflate the features of the media (language) to be those of the content (meaning) and the other way around.

2.2.1 Psychology of Language vs. Psychology through Language

A fundamental distinction that needs to be drawn here is between the *psychology of language,* and the *psychology through language*. The former relates to those elements of psychology which make possible the manifestation of language-like abilities. On the other hand, psychology through language is concerned with using language as merely a tool to discover more general features of human cognition. When we talk of syntax, lexicon, speech production, etc., we are concerned with features of language that are implemented in mind but are unique to language and hence fall under the category of the psychology of language. On the other hand, conceptual knowledge and meaning are of importance not just in language and communication but also planning and action. To use language to theorize about these would be psychology through language. Unfortunately, this distinction is not often made in psycho-linguistics research.

It is also possible that cognitive elements earlier thought to lie solely in the domain of language are found to actually be more general features present in non-linguistic cognition as well. For example, generativity—the ability to produce an infinite number of complexes from

finite means—was thought to be a feature unique to language but has now been established to be prevalent in many non-linguistic abilities as well. Thus, when using language to study meaning, one must be careful about the kind of attributions one makes and requires us to separate the features of language from those of meaning. As I discuss later, models that conflate the two are open to much criticism. It is thus necessary to look at the current understanding of linguistics so as to delineate the parts which are relevant to develop a theory of meaning, from those above and beyond it.

2.2.2 Compositionality and Generativity

The notion of compositionality is a powerful one and has frequently resulted in dramatic development in the understanding of the various fields the phenomenon of which it has been deployed to explain. In its most generic form, compositionality aims to describe a complex entity/event as arising out of different parts coming together in different ways. This goes as much for material sciences such as physics, chemistry, and biology as for more abstract semantics and mathematics. In semantics, the principle of compositionality claims that the meaning of a complex sentence is a function of the meanings of its parts together, and the manner in which these parts were synthesized. The parts aren't just important for what they individually do, but also because in coming together they create, and thus explain, something else. "If you have a compositional theory of a certain kind of system, then, you do not need to theorize anew for each instance of that system" (Strevens, 2017). Generative systems which have a property of being recursive expand the compositionality of the system. That is, by being able to use composed outputs to re-compose more complex structures, recursion expands the power of

compositionality resulting in the possibility of infinite output from finite means—basically compositionality on steroids. It was this generative compositional nature of language which Chomsky sought to explain through syntax.

For Chomsky, this generative ability derives from a generative syntax on which semantic and phonological structures of language depend (Partee, 2014). Jackendoff suggests that syntax, semantics, and phonology each have their own generative systems which are communicating in parallel with each other; and furthermore that actions are also privileged, like language, in regards to access to recursion-like functions (Jackendoff, 2007b). Levelt (1993) had earlier proposed that communicative meaning/intentions can vary in infinite ways and are produced by a processing system he had called the conceptualizer. Jackendoff described semantic relationships as "complex, embedded, recursive and multi-dimensional". Snyder et al. in their 2004 paper, propose that object attributes that are stored in the brain result in concepts. When that happens, we become aware of the concept but unconscious of the attributes underlying the concept. This continues as we form meta-concepts (grouping of concepts), and become unconscious of the concept while we experience awareness of the meta-concept (Snyder, Bossomaier, & Mitchell, 2004). This is where, according to me, the real power of representations comes from. Existing representations allow for coding new ones and leads to ever-increasing complexity. This increasing complexity has occurred even in the evolution of artificial computational languages and is a general good practice in coding. When machine language used to program microprocessors became too cumbersome to handle the kind of complexity newer algorithm demanded, it was used to create higher languages which were then used to create

even more languages. The reason why this was possible was because the complex objects of each language became base objects for the next. The algorithm that multiplied binary numbers in machine language became a simple one-word command for the next one. In fact, coders of this higher languages usually are not even familiar with how these commands are executed in a lower language. The complex processes of one level become the constitutive bricks for the next level.

The issue is no longer just what is compositional and generative in humans, but rather, how is it that much of what we do is. We can thus draw compositional complexity and generativity as significant constraints for a theory of meaning.

2.2.3 Words vs Concepts Distinction

It is relatively uncontroversial that word meaning (i.e., lexical-semantics) is grounded in conceptual knowledge. More difficult is the question if the two are distinguishable from each other (Vinson, 2009). The units of natural languages are words; and many models of meaning work under the paradigm of treating concepts to be represented by words. As a result, 'virtually all computational models of semantics, from the early semantic network, to electronic thesauri such as WordNet and common-sense database like Cyc, all the way to purely statistic approaches like LSA…model relations between words"(Roy, 2008). These models thus take natural language corpora as their inputs. Given the internet boom, and a large amount of unstructured data in the form of text on the internet, it has been highly lucrative to create models which can process such data to extract meaningful insights and inferences from it in an automated manner. An accurate extraction of meaning from natural language is the Holy Grail for most tech-giants of the digital world.

That said, word and concept are not the same things. It is known that speakers of a language know more concepts than words (Murphy, 2004). For example, the concept of 'a confused left-right maneuvering when we almost walk into someone and end up blocking each other's path repeatedly' is familiar to most English speakers and yet English does not have a word for it. And it is possible that one can invent a word for it and make its usage commonplace. There thus seems to be thus certain arbitrariness relating words and concepts instead of a direct mapping. There is also psychological dissociation between the two evidenced from cases of patients suffering from aphasia with a severe naming disorder but without loss of the conceptual knowledge of the objects (Chertkow, Bub, Deaudon, & Whitehead, 1997; Chomsky, 1980). The difficulty to separate words from concepts goes deep. An example that Chalmers et al., (1992) use is that of the word *Hard* which has many flavors including *alcoholic*, *severe*, *difficult*, *callous*, *rigid*, *industrious*, *intense*, etc*.* Despite the different readings, there is a feeling that the overarching concept is still the same. But in German, the modifier word used for *a hard substance* is *hart*, for a *hard problem* is *schwierig*, for a *hard blow/hit* is *schwer,* for *a hard life, in general,* is *mühsam*. A German speaker would not see as much unity behind these modifiers as an English speaker sees behind various uses of *hard*. And it is not like these are all narrower version of the concept *hard.* In fact, these German words have multiple broad flavors of their own. Thus, the unity we see behind words in the form of an overarching abstracted concept is a false unity.

A significantly radical proposal of Jackendoff is a reformed conceptualization of the entity "word". He argues for conceptualizing words as more than "an arbitrary association of a chunk of phonology and a chunk of conceptual structure, stored in speakers' long-term memory (the

lexicon).”; but rather as complex items, each with partly idiosyncratic properties such as syntactic markings (verb, preposition etc.), grammatically encoded arguments (agent, theme, path), etc. which “critically govern how the word enters into the recursive components of grammar” along with the conceptual and phonological elements (Jackendoff, 2007b). Thus words, as imagined by him, contain rules from generative semantics, generative phonology, and generative syntax combined together.

The models of semantics which use natural language corpora are thus actually not even models of semantics, but of complex linguistic entities—words, and combinatorial relations between them which are different from conceptual combinatorics. Hence, it does not make sense to use natural language expressions to directly encode conceptual relations. The rules of natural language carry content through linearly expressed phonetic symbols; and thus, constitute rules for transformation of complex, embedded conceptual representation into linear expressions. As a result, we have many sentences which are ambiguous in their meaning and different sentences which have the same meaning, idiosyncratic features such as “a-an” articles, rules of merging words, and sentences which mean something entirely different than the meaning of their components, and so on. Given the dissociation of between language and conceptual meaning expressed through the language, it does not make sense to model meaning by directly using natural language corpora which have linguistic structures enmeshed in them. Modeling conceptual information requires a move from natural language to a deeper meaning structure. We thus have a constraint over the kind of inputs which can be used to model meaning.

2.3 Psychological and Psycho-linguistics constraints

"Felicitous advances in linguistics presuppose an advanced psychology."

-Heymann Steinthal

It was with above conviction that Heymann Steinthal worked to develop a truly original proposition that meaning lies in the unconscious and that language is only a way to consciously access this meaning. As discussed earlier, linguistic behavior has been used both as a subject to be studied, and as a tool to study cognition. Research on psycholinguistic phenomena like priming through words gives us an insight into how concepts interact with each other. What Steinthal aimed to create was a psychology of language which, first and foremost, treats language as merely a surface phenomenon that allows access to unconscious meaning. For this, Steinthal attempts to combine Johann Friedrich Herbart's mathematical theory of thought with Hermann Lotze's theory of the reflex (unconscious) action (Levelt, 2013). Interestingly, Herbart occupied the chair formerly held by Kant at Königsberg and largely disagreed with Kant's idea of innate categories. Instead, he was an empiricist and derived a mathematical mechanics of mental representations which was highly associationist. In these mechanics, the stimuli and representations can interact in following ways:

- **Fusion**: Stimulus of an object whose representation already exists does not create another representation, rather just evokes previous representation. A+A = A.
- **Entanglement**: Different representations with shared properties will fuse at the commonalities, and repel at the differences. Such entanglement can cause the formation of complex ideas. For example, the representations of a car and a scooter are fused in

their commonalities (both vehicles, both have wheels) but separated by the difference (different sizes, number of wheels, driving styles)

- **Association**: The process by which a representation gains energy from another representation which is active. This can be seen as resulting in a chain of thoughts, one leading to another.
- **Apperception**: New representations are created by forming linkages to one's unique set of existent representations. This growing mass of representations is called the apperceptive mass and can be treated as individual's unique epistemology.

This kind of mechanics is exactly reflected in a modeling paradigm used for knowledge storage known as Vector Space Architecture which we will discuss later. Interestingly, modern studies on the neural perspective of the brain has found evidence for these interactions in form of activation, co-formation, and spreading of electrical signals in neural assemblies as a manner in which concepts are implemented in the brain and interact with each other (Bauer & Just, 2015).

While the linguistic approaches allow a glimpse into the combinatoric nature of meaning, the psychology of mental representation gives us the foundational set of constraints for the architecture over which this combinatorics is implemented. Some of the few features of mental representations that fall out of such kinds of dynamics are those of gradience of meaning and spreading of activation to related concepts (semantic priming). Thus, the architecture for the models of meaning need to be selected such that they allow for such cognitive phenomenon to emerge. Later in thesis, I propose using Vector-space models as a way of modelling this architecture because it satisfies these constraints.

2.4 Perceptual Grounding

Most of the language-based models of meaning are criticized on grounds of lack of embodiment of meaning resulting in AAA—arbitrary, abstract, amodal symbol systems (de Vega, Glenberg, & Graesser, 2008). What this implies is that these models ultimately require a human interpreter to make sense of them. For example, there is nothing in the symbols representing the color *red* which entails the *redness.* The same is true for concepts no matter which modality they are from; there is something inexplicable about sweetness, hotness, anger etc. which in order to be described needs to resort to how it feels to taste something sweet, to feel hotness and to feel angry. Searle's Chinese room experiment brings out exactly this issue. Searle shows that mere computation through symbols is not sufficient to account for cognitive experiences amounting to meaning. It is because they approach conceptual knowledge from a linguistic and purely computational point of view that these models are divorced from perceptual and motor modalities which contribute to both their basis and use. On the other hand, the *Embodied Cognition Framework* claims that "semantic knowledge is grounded in sensorimotor systems, that are automatically engaged during online conceptual processing, re-enacting modality-specific patterns" (Gainotti, 2011).

As such then it is essential not to locate meaning merely in amodal relations; but rather, to create architectures which are capable of drawing meaning from multiple modalities. This is not a novel requirement that is put forth for the first time in this thesis, but the reason to bring them here is to make problematic the separation of amodal symbolic models of meaning and theories of grounded meaning. This separation is symptomatic not of intent but rather of lack of

frameworks which can combine the two domains of human cognition—the abstract linguistic thought and grounded, embodied experience into a holistic conceptual space. This was the kind of bridge that Kant's critical philosophy attempted to build (see section 6.2.1).

In his critical work, Kant is concerned with exactly these kinds of rules when he talks about his doctrine of *transcendental aesthetics* and *transcendental deduction.* For Kant, the inquiry was of the link between senses and concepts which explains and justifies how we can refer to the objects of senses using our concepts.

2.5 Selection of Primitives

Compositionality and generativity make it possible to construct a great deal from a small set of primitives—infinite from finite. Primitives here can refer to either atomic units which are bound together to create complexes, or the set of rules through which the binding occurs. It is both the diversity of components and of methods of binding which make a compositional system productive. Any composition-based theory of meaning would thus need:

i. a theory of conceptual components (Matter)
ii. a theory of binding of these components (Form)

The choice of these primitives plays a key role in determining the nature of the output of these systems. This leads to a critical issue. The problem is similar to what linguists encountered while creating grammatical rules for different languages. Some of the model-languages were too restrictive to generate all of the possible sentences in the language, while others were too liberal and would generate all the grammatically correct sentences and then some more which were not grammatically correct. The models thus need to be appropriately constrained through a right

selection of primitive rules, such that they do not over-generate or under-generate (Isac & Reiss, 2008).

This issue has been found in semantics as well. In a 1994 paper, Zadrozny (1994) shows that "compositionality is not a strong constraint on a semantic theory" by proving a theorem that states that any semantics can be encoded as compositional. This for Zadronzy makes compositionality a vacuous descriptor of semantics. What that means is that mere compositionality is insufficient as a constraint over models of meaning since it does not allow us to locate a unique model for the same. The solution they propose is that "the meaning functions should be systematic, i.e., non-arbitrary". Dever (1999) encourages consideration of what the effect of requiring constraints on various parts of the language would be. In order to have a compositional and generative model of meaning which produces patterns observed in human thought, and only such patterns and no other, we see a necessity of constraints over the primitives. But Berwick (1989) argues that the selection of semantic primitives is a "hazardous game". How does one validate the selection? This issue is faced by any compositional theory, and thus a compositional model of meaning is not just what the primitives in conceptual memory are but also the principle through which these are derived.

Another issue here is of the source of primitives. Are the conceptual primitives innate, i.e., are we born with them, or are they acquired? Empiricist approaches claim that even our most basic concepts are derived from the world. This approach is often useful in constraining the complexity of the system and prevents theories which explain away complex behavior by presupposing the complexity in the system itself. But research in the fields of child development

suggests some minimal innate structures are necessary to explain the rapid skill and knowledge acquisition by children despite impoverished stimulus. This is what Chomsky called Plato's Problem (Chomsky, 1988) which was expressed by Bertrand Russell as "How comes it that human beings, whose contacts with the world are brief and personal and limited, are nevertheless able to know as much as they do know?"(Russell, 1948).

Thus, there is a reason to believe that the mind may have innate structures that provide a scaffold for acquiring knowledge of the world. According to Kant, this innate structure is not merely a power to extract patterns from nature but is "legislative" of the patterns we can experience in nature, determining the ways in which we can conceptualize the world. This is especially important because unlike other attempts at proposing categorial primitives, Kant's primitives are relational.

2.5.1 Relational Primitives

As discussed earlier, there are two ways to constrain a compositional system—either through its components, or the way they are bound. Controlling the methods of binding in a model is to control/fix the ways in which it is generative. On the other hand, controlling the primitive components would determine the kind of objects that we can bind.

Much progress has been made in terms of robots and AI systems like Shakey (Nilsson, 1984) which can process commands and act on them in the real world in a systematic way by being sensitive to the features in the world and mapping them to the commands. Given this, one might ask what gap is left between human cognition and AI systems which show both computational and embodied cognition. The answer to that lies in a key difficulty with such

systems which is that their conceptual system is programmed into them. Unlike humans, they lack an ability to construct new schemas and concepts apart from what has been given to them. These systems have primitives in form of pre-coded concepts, but without relational primitives, they lack complex generativity which allows them to create new schemas. The issue is later discussed in the form of a critique of first-order predicate logic with respect to forming mental representation. The logic assumed objects as given, while the mentalistic models must essentially construct their objects. A more detailed discussion occurs in sections 4.5.1 and 6.3.1.

Thus, primitives in the form of relations which allow for the creation of new concepts are necessary for autonomous generation of a large conceptual repertoire. The model of the conceptual system proposed in this thesis uses a constructive architecture which binds relations between concepts and I propose using Kant's categories to expand the ways in which these bindings take place in this architecture.

2.6 Chapter Conclusion

Constraints for modeling meaning spread over a large ambit of research including:

i. formal linguistic constraints like compositionality and generativity to psychological ones regarding knowledge storage and concept activation.
ii. selection of primitive- content and relational.
iii. with regards to using amodal symbols or embodied perceptions as inputs
iv. and finally neurally plausibility.

Keeping these in mind, I will next discuss some of the models of meaning.

Chapter 3: Semantic Models

3.0 Chapter Overview

In this chapter, I will discuss some models which have been created under these paradigms. and conclude that none of these fulfill all of the constraints identified in the previous chapter. I locate the issue in a lack of sufficiently general architectures for human cognition.

3.1 General Distinction Between Semantic Models

Many models of meaning have been proposed in the literature, some of which are components of an overarching system to model language. These approaches can be roughly divided into *logical semantics* and *distributional*/*statistical semantics*. *Logical semantics* models concepts as symbolic entities bound in set-theoretic, rule-based relationships with each other. The issue with the logic-based approaches to semantics is that "representations in terms of logical formulas are not well suited to modeling similarity quantitatively" (Mitchell & Lapata, 2008) and thus do not display a graded sense of meaning which is a mark of human meaningfulness. Thus, while they are compositional and generative in nature, they do not by themselves support the artifacts around human experiences of meaning. These approaches also assume a significant innate structure in the mind. Empiricists, on the other hand, favor approaches where meaning is derived from empirical sources (perception), and language is a platonic entity. These came to be referred as distributional-*statistical* approaches. I will discuss next the different kinds of models which fall under these categories.

3.2 Classical Models

The theory of meaning behind classical models was that the meaning of a concept is its definition; i.e., a set of features describing all examples which fall under the concept. The meaning is also assumed to be compositional. This view dominated theory of meaning through much of the 20th century. The first kind of models that resulted from this approach were created with the aim of developing structures which allow for first-order logic relationships. Some of these models, inspired by the work of Chomsky, acknowledge an extra layer of linguistics machinery which transforms intentional meaning into natural language token or surface forms (Jackendoff, 1983; Katz & Fodor, 1963). They thus located their research in finding ways to extract this logical form. Others assumed that the formal languages, inherited from the Frege-Russell-Tarski tradition which had given rise to the current model-theoretical practice in logical semantic theory, to be same as the logical form of natural language and attempted to create semantic structures which relate concepts in this manner. Meaning is described as concepts related symbolically in set-theoretic rule-based relationships with each other. These include defining operators such as conjunction and disjunction, quantifiers like *some* and *all*, and giving concepts functional and thematic roles. The meaning is exactly expressed in its entirety through formal systems like lambda calculus. These models thus presume organizational and functional primitives through which combinatorial systems of meaning arises.

The pride of these models is their inferential power and lack of ambiguity. The logical form demarcated definitions exactly and precisely. The criticism of these models came from two grounds—firstly, Wittgenstein points at the difficulty of definitions to truly encompass all

meanings under a concept. For example, the concept *Game*. The many examples of games have no set of core features which is true for all of them. The same is true for most concepts. Hence, these models frequently fail due to their inability to account for the blurriness of meaning and its sensitivity to context. They are thus unable to account for gradience of meaning and fail to duplicate the psychological constraints around meaning comprehension in humans. Secondly, these models are also purely formal and computational, and thus provide no explanation of either knowledge acquisition or storage. These are deferred to perception and general cognition without any interfaces with the current theories of either phenomenon.

3.3 Feature List models

Feature-List-based models take the compositional approach of classical models and attempt to describe all concepts through few base concepts or features (E. E. Smith & Medin, 1981). It has its basis in Aristotelian categories, and the idea is to identify these base concepts and then describe all other concepts as a composition of these. These accounts assume the existence of definitions for concepts in the classical sense (core features), but also propose another set of (non-necessary) features to reflect information that is not necessarily part of the definition itself, but instead, properties of some, but not all, exemplars of a category. Although core features will always be relevant (because they are common to all members of a category), non-necessary features would be used for identification procedures, as they are often more accessible than the core features (Smith & Medin, 1981). This allows for certain fuzziness in how concepts are defined. Many of these models get their list of features based on common descriptors used by the participants to describe concepts. The issue with these models is that

while they describe concepts based on components, they do not say anything about how these features come together to form the concepts. In other words, they describe concepts as bags containing certain features without a generative theory which allows for novel combinations of concepts to form new concepts.

Due to their emphasis on definition, both feature list models and classical models are unable to address the issue raised by Donnellan (1972) regarding the ability of descriptive accounts of names (individual or generic) to always give a unique and intended set of referents.

3.4 Semantic Differential Models

Given the difficulty in definitional approaches to meaning to be able to address the fuzziness of the meaning of concepts, a relational approach was taken. Instead of focusing on one concept, this approach looks at the ways in which concepts differ from each other. Hence it is called *differential* semantics. This was a more psychological approach bringing the *mental-ness* of meaning into the forefront. The models ask participants to rate the difference between concepts based on different scales (like good-bad, strong-weak, active-passive, adequate-inadequate, real-imaginary etc.) with neutral as an option for all. The rating difference represents the psychological distance in the meaning of two concepts along the dimensions constituted by the scales. Some models, instead of using such domain-general scales (like good-bad), use domain-specific scales which are particular to the topic the words relate to. An intuitive analogy would be to compare and rate the experience of two flavors based on scales of saltiness, sweetness, sourness, texture etc. The scales are domain specific in this case.

This approach falls under the category of distributional models. The biggest issue with this approach is that because the similarity between concepts is measured in a graded sense across only fixed categories of relations rather than specific defining characteristics, it is difficult to individuate concepts. For example, the concept of a bird is more than just a list of how it varies along a few fixed general dimensions (like active-passive, animate-inanimate etc.).

3.5 Semantic Network Models

In *Semantic network models,* concepts are stored as nodes in a graph, while the edges represent semantic relationships between them. Meaning is determined in terms of connections with other concepts. Semantic network models are thus distributional in nature. The models are usually designed to be sensitive to special relations between words, and thus have to be parsed a priori by the modeler. For example, Collins & Quillian (1969) propose two kinds of relations in their model—property and sub-set. Hence, their model captures inheritance-based relations well. These models are a mix of formal models and relational models in that meaning arises from relations between concepts, but the way information is structured (through set-subset system, e.g., to relate *cat* as subset of *animal*), and the selection of relationships (such as *can*, *has;* for example *animal* can eat, *cat* has claws) allows for some inheritance-related logical inference. Perhaps the most extensive model which implements distinct representational themes is Wordnet (Miller & Fellbaum, 1991), a network model of the representations of a large number of nouns, verbs, and adjectives in English. In Wordnet, “nouns, adjectives, and verbs each have their own semantic relations and their own organization determined by the role they must play in the construction of linguistic messages”. That said, the models have limited logical forms and

use primitives borrowed from syntactic theory instead of a theory of primitives based on conceptual organization. They thus do not represent conceptual meaning, but instead are representations of the lexicon; words as Jackendoff envisages them –complex syntactic and semantic entities.

3.6 Semantic Space Models

Somewhat similar to the Semantic network models are the *semantic space models* which work directly off the natural language corpora. They presume little or no pre-requisite innate structure whether syntactic or semantic. These models are founded on Behaghel's First Law—elements that belong close together intellectually will also be placed close together in sentences (Behaghel, 1932). Such models tend to represent concepts as word-nodes which get their meaning through co-occurrence statistics and order-based proximity relationships with other words. For example, related words such as *wheel*, *gas*, *transport* etc. would co-occur with the word *car* in sentences. These would also co-occur with *bus* and result in close proximity between nodes of *car* and *bus*. In fact, the 1997 paper by Landauer & Dumais, which introduced Latent Space Analysis as one of the first statistical semantic approaches, claimed to have solved the aforementioned Plato's Problem. Semantic space models are distributive and statistical in nature since the meaning of the concept is situated in the concepts occurring in its immediate vicinity, and the proximities are computed statistically.

A critical issue of a lack of logical structure in these models leads to a lack of inference ability with regards to analogy, inheritance of attributes, etc. Semantic space models output unorganized clusters of connected words/concepts with no inferential power (Glenberg &

Mehta, 2012). These models are not compositional in the sense of having concepts constructed via recursive integration. Rather, concepts are said to be composed of other concepts only in so much as that concepts (words) which co-occur are clustered together thus contributing to each other's meaning. Also, most of these models work with natural language corpora without accounting for syntactic structures. They are thus unable to account for long-distance dependencies between related words in a sentence and would thus presume unrelated co-occurring words as related.

3.7 Chapter Conclusion

What is apparent in the discussion of these models is that no one of these fulfills the set of criteria—compositionality and generativity, inferential abilities, compatibility for neural implementation, perceptual grounding, psychological constraints like graded meaning, etc., to model the full breadth of how humans experience meaning. A crucial gap which is being felt by cognitive scientists across different domains of the research is that of a lack of sufficiently general architectures that can implement multiple kinds of mental representations and synthesize a singular whole from them in a manner that satisfies all these above criteria. Jackendoff (2017) echoes this when he says "…we do base abstract reasoning on principles of physical experience, we do learn on the basis of instances, and statistical and probabilistic factors are involved in our perception and behavior. But to my taste, something important is missing from all of these approaches…an explicit and sufficiently general theory of the mental structures being computed and learned." This requirement directly speaks to the constraints we discussed earlier regarding the general psychology of mental representations. We discussed the Herbartian mathematics of

mental representation to highlight the kind of mechanisms which such a general theory of mental structures must have. A possible solution could be in form of Vector Symbolic Architecture (VSA) which is a connectionist network architecture specifically developed to create systems for symbol instantiation and manipulation.

Chapter 4: Vector Symbolic Architecture (VSA)

4.0 Chapter Overview

The chapter introduces and explains Vector Space Architecture and the models based on it. Many of the constraints posed by the psychology of mental representations that we discussed get automatically satisfied by the sub-symbolic aspects of VSA, while it is still able to implement symbols (Kelly & West, 2012). I then proceed to discuss a model created by Rutledge-Taylor, Kelly, West, & Pyke (2014) to store conceptual knowledge. I discuss the advantages of VSA platforms that have been used to encode both semantic information and spatial and perceptual data, and hence conclude that VSA is promising with respect to the issues of grounding. I end the chapter by highlighting the strengths of the architecture and the issues with current models of VSA which are resolved in the later chapters of this work.

4.1 Vector Space

A vector space is defined by a set of perpendicular axes such that all the points can be defined through their location corresponding to these axes. These axes represent the degrees of freedom across which a point can move. For example, in a 2-dimensional vector space, we have two orthogonal (90-degree angle between them) axes—***x*** and ***y***. A vector in this 2-D vector space can be understood as a line segment from the center of the coordinate system (defined as [0,0]) to a point defined by values of its distance from the two axes. For example, a vector [1,2] is at a distance of 1 units from the ***x***-axis and 2 units from the ***y***-axis. In the figure below, you can see three different vectors plotted on a 3-dimensional system, i.e., a vector space with three mutually

orthogonal axes—***x***, ***y***, and ***z***. These vectors are represented by the arrows from the center (represented by the point [0,0,0]) to the respective points— [7,7,8], [-4,7,3] and [-7,-9,7].

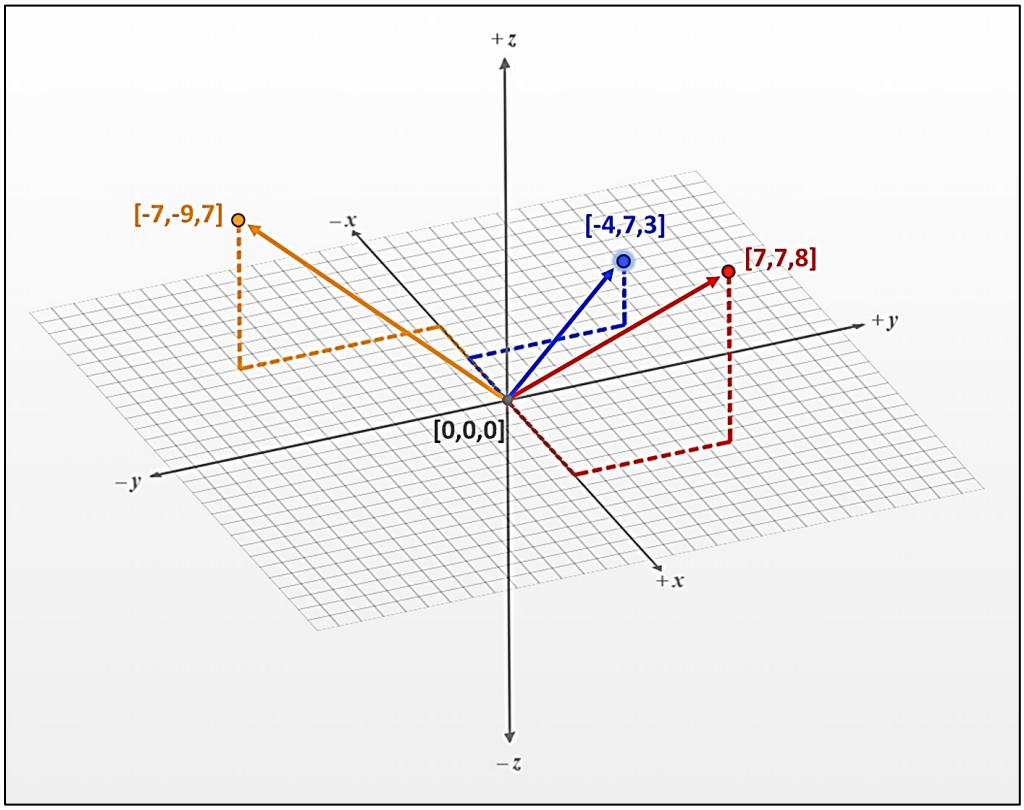

Figure 3: Examples of 3-D Vectors

Features of vectors and vector space, such as the angles between the vectors, are important with respect to deploying VSAs to store relations between concepts. For example, the cosine between the vectors is used as an indicator for similarity between concepts. The Cosine between the two vectors—say A and B, equals the dot product of A and B divided by the product of their magnitude (see Equation 1). Taking the inverse of this cosine value gives us the angle between the two vectors. The cosine value can vary from 0 to 1. If the two vectors are orthogonal (90 degrees), their cosine value is 0 whereas if the angle between them is 0 degrees, their cosine

value is 1. This makes cosine a very useful property. We will discuss properties like these in more detail in later sections.

$$\text{Cosine (A,B)} = \frac{\text{Dot Product (A,B)}}{\text{magnitude(A) x magnitude (B)}}$$

Equation 1: Cosine between vectors

4.2 Storing information in Vector Space

In VSA, the concepts are represented by unique vectors in a high-dimensional space (also called *hyperspace*) in which they are instantiated. A hyperspace is like a Cartesian coordinate system, but instead of a conventional coordinate system with a small number of dimensions (usually labeled x, y, and z), it has a large number of dimensions usually exceeding 64. Due to their instantiation in a spatially governed system, VSAs are geometric systems. The relationship between any two concepts is stored by moving the corresponding vectors through the high-dimensional space in which they are instantiated. This is achieved through vector addition, also called the *superposition* (indicated by +). Superposition of a vector (say B) onto another vector (say A) moves A close to the vector B. As a result, the angle between them gets reduced (Figure 4). This results in the behavior that related concepts have an angle less than 90 degrees to each other, and thus, the smaller the angle between any two vectors, the more similar they are. Because the similarity between the concepts can vary only from 0% to 100%, and the magnitude of the cosine any two vectors can only vary from 0 to 1, cosine becomes a useful way to indicate the similarity.

Figure 4: Angle reduction to reflect the similarity of concepts

Once the vectors are initialized, as the system is given more and more relationships, vectors that contribute to each other's meaning cluster together. More importantly, this causes the vectors (say X and Y) that share relationships with the same other vectors (say P, Q, R) will also tend to be close to each other. This happens even if no relationship between them had ever been explicitly stored. As a result, VSA results in a global storage of information.

4.3 Compositionality in VSA

In addition to Superposition, vectors in VSA can also be bound to indicate a new concept. Binding is usually represented as ⊗. A *binding* operation uses two vectors to generate a new vector from which the earlier two vectors can be later extracted through a release operation represented as ⊘. Good binding operators thus obey at least the following rules:

i. For two vectors **A** and **B**, **A** ⊗ **B** is a new vector, i.e., usually not similar to either **A** or **B**.
ii. Binding has an inverse or at least an approximate inverse operator ⊘, called the release operator, such that (**A** ⊘ (**A** ⊗ **B**)) ≈ **B**.

Vectors can be bound recursively to create embedded complex structures representing a combination of those vectors. This is useful in the situations where we do not want to move the original concepts close to each other but want to store a relation which binds them. The release operator ensures that, given a bound product and one of its constituents, we can recover the other constituent with reasonable accuracy.

In the models discussed and developed in this thesis, a mathematical operation called the *circular convolution* (indicated by *) is used to bind the vectors together. The advantage of circular correlation over some of the other binding operations like the *tensor product* is that operations like the *tensor products* result in increased dimensionality after every operation. Circular convolution, on the other hand, outputs a vector of the same dimensionality as the input vectors. The release operator corresponding to the convolution is a mathematical operation called the *circular correlation* which gives us an approximate of the inverse of the convolution. For mathematical details of these operations see Plate (1995).

VSA models thus exhibit both compositionality and generativity. These features made VSA a promising architecture for many connectionist researchers who wanted to respond to the doubts raised by many linguists at the ability of the connectionist approach to model the core features of compositionality and generativity found in language (Gayler, 2003). VSAs have been used to create semantic space models (Jones & Mewhort, 2007) which extract the meaning of words from co-occurrence statistics in unparsed natural language corpora. But such statistical models of semantics have faced the criticism that word co-occurrence and order are insufficient to account for the linguistic structure in which the meaning is encoded (Jackendoff, 2007a). Such

criticism pertains to our earlier distinction between word and concepts. To bypass this, Rutledge-Taylor, Kelly, West, & Pyke (2014) used the VSA architecture to model conceptual memory by directly encoding the related concepts.

4.4 Dynamically Structured Holographic Memory

Rutledge-Taylor, Kelly, West, & Pyke (2014) used BEAGLE (which is a VSA-based language model by Jones & Mewhort, 2007) as a general-purpose model of human memory, instead of just as a model of language acquisition. The system is called Dynamically Structured Holographic Memory (DSHM). Kelly & Reitter, (2017) integrated DSHM into ACT-R (Anderson & Lebiere, 1998), a widely used cognitive architecture that can model diverse aspects of cognition. The subsequent system was called HDM—Holographic Declarative Memory.

Both HDM and DSHM systems use the BEAGLE's VSA architecture to directly encode conceptual relationships instead of using natural language corpora. As a result, although they need *a priori* assembly of information, they escape the criticism faced by models which attempt to directly model meaning relations from natural language corpora. These models (like BEAGLE), have two vectors for each concept:

i. an environmental vector ($\mathbf{E_c}$) which represents the percept of the concept, and can thus be understood as its referent.
ii. a memory vector ($\mathbf{M_c}$) which encodes the concept's relationship with other concepts. A concept has meaning as a result of these relationships, and thus $\mathbf{M_c}$ can be understood as the sense of the concept.

The reference of the concept (environment vector or $\mathbf{E_c}$) is fixed and plays the role of the identity of the concept in the system and is used to create the vectors which are superimposed onto the memory vector to store a relationship. The sense of a concept (memory vector or $\boldsymbol{M_c}$) is the vector which is moved (through superimposition) to stores all the relationships which the concept has with other concepts. The initial value of $\boldsymbol{M_c}$ is the same as $\boldsymbol{E_c}$.

Let's say I want to encode the information about a '*red door'.* First, the environmental and memory vectors for the two concepts involved are initialized (Table 1).

Table 1: Memory Chunk

Concept	Environmental vector (E_c)	Memory vector (M_c)
door	$\mathbf{E_{door}}$	$\mathbf{M_{door}}$
red	$\mathbf{E_{red}}$	$\mathbf{M_{red}}$

Multiple paradigms exist to encode this information. The first is a simple vector addition or direct superimposition. In this method, the memory vectors of both the concepts are changed by adding the environmental vectors of the other concept into them (see Equation 2). This results in a (decreased angle) increased similarity between the two concepts (Figure 5).

$$\textit{New } \mathbf{M_{door}} = \text{Old } \mathbf{M_{door}} + \mathbf{E_{red}}$$

$$\text{New } \mathbf{M_{red}} = \text{Old } \mathbf{M_{red}} + \mathbf{E_{door}}$$

Equation 2: Storing relationship through simple addition

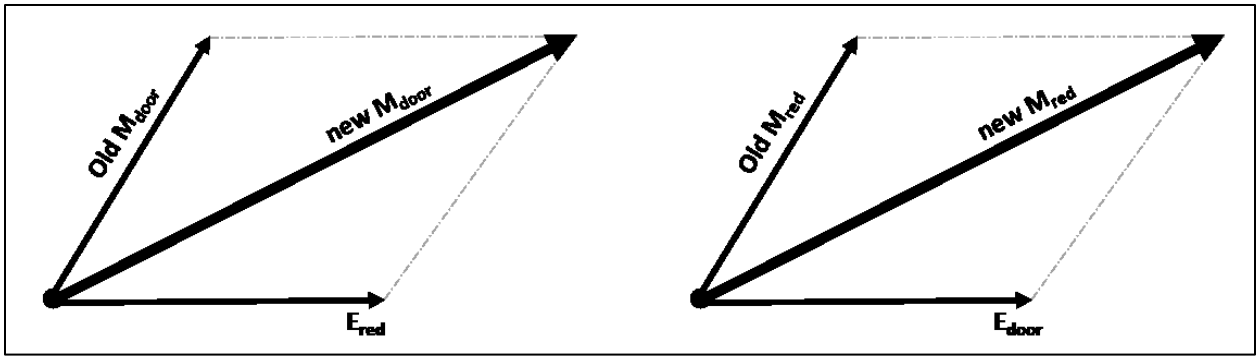

Figure 5: Encoding using simple addition

But it is not always desirable to bring the concepts close to each other when the aim is for them to merely relate to each other. We might not want the concepts of *red* and *door* to share meaning, but merely represent the idea of them being related—of a *door being red.* Hence, in DSHM, all encodings are done through the placeholder vector—Φ. Placeholder vector is a randomly generated vector which once generated, stands universally for the notion of *self* throughout all encodings. This vector is bound with the environmental vector of a concept through the binding operation ($\Phi^* \mathbf{E}_{concept}$) and added to the memory vectors of all concepts which relate to that concept. For example, $\Phi^* \mathbf{E}_{red}$ represents "all things red". The information of *red door* will be thus encoded in the memory vectors as follows:

$$\text{New } \mathbf{M}_{door} = \text{Old } \mathbf{M}_{door} + \Phi^* \mathbf{E}_{red}$$

$$\text{New } \mathbf{M}_{red} = \text{Old } \mathbf{M}_{red} + \Phi^* \mathbf{E}_{door}$$

Equation 3: Storing Relationships via Convolution

This moves $\mathbf{M}_{door}$ towards the vector $\Phi^* \mathbf{E}_{red}$, i.e., towards all things with the property of being *red*, and moves $\mathbf{M}_{red}$ towards vector ($\Phi^* \mathbf{E}_{door}$), i.e., towards all things that are *door*. Hence the steps involved in encoding are as follows:

i. bind the placeholder vector (Φ) with the environmental vectors to create a new vector through convolution (*);
ii. add the new vector to the memory vectors using superposition (+), thereby moving it to represent the addition of the new relationships to memory.

As more relationships are added to the memory vector, the memory vector of the concept becomes richer. At the same time, because convolution is reversible, the information can be retrieved through a cue. Because of its Vector Space architecture, DSHM automatically spreads the impact of the change in information across all conceptual relationships and allows for similarity between concepts to exist as a gradient (a varying angle). These features allow for storage of information in a manner that mimics similarity-based aspects of human memory.

4.5 Assessment of VSA

The vector-based distributional models made an impact in cognitive science through models of computational linguistics which began with Smolensky (1990). It has since inspired many more (Burgess & Lund, 1997; Landauer & Dumais, 1997; Padó & Lapata, 2003). These models "which automatically induce word meaning representations from naturally occurring textual data, are a success story of computational linguistics" (Baroni, 2013). The architecture has been used to model meanings of not just words but also sentences (Jones & Mewhort, 2007; Kachergis, Cox, & Jones, 2011; J Mitchell & Lapata, 2010). In models like DSHM and HDM, VSA is used for conceptual memory representation and performs well with regard to psycholinguistics constraints. But VSA models face criticism due to following two reasons:

i. the lack of an organizational schema in VSA leads to the issue of bag-of-words/concepts and no inferential abilities in these models.

ii. insufficient understanding of the mathematical aspects of VSAs leads to lack of methodologies required to systematically build models that address the above issue.

In this section, I will first discuss these two issues and then make a case for why we should not give up on VSA as a basic framework to develop models of meaning.

4.5.1 Bag of Words/Concepts

The primary issue that arises with all models of VSA is the bag-of-words/bag-of-concepts issue. The issue concerns lack of organizational structure in these models. While these models "…indicated a degree of similarity between two items, [but] not any particular relationship since the vectors are inherently 'non-meaningful'. The representations make no commitment to a particular set of features or theory of meaning, although the vector representations imply a certain degree of relatedness in order to model cognitive effects" (Burgess & Lund, 2000). There are other models like Padó & Lapata (2003) which use parsed natural language data to store co-occurrence based on syntactic structures instead of sequential order. While they seem to be an improvement over word-sequence-based model, the compositions are still non-meaningful without any systematic structure to organize them. The only change is in the selection of the words that are bound together. Due to these lacking, the models have no inferential ability apart from how similar two things are without being able to shed light on the nature of the similarity. The issue here is that these models make the same assumption which empiricists made with respect to knowledge acquisition from experience. The empiricist assumed that we pick up knowledge from the world as is, and the only thing that mind does is store and reproduce this knowledge. In a similar tone of thought, these linguistic-computational models assume that meaning is present in natural language corpora, in how the words co-occur or syntactically relate.

Jones & Mewhort, (2007) state this exactly when they say that their BEAGLE model "builds a semantic space representation of meaning and word order directly from statistical redundancies in language". What these natural language VSA-based models are thus attempting is the extraction and storage of meaning with an assumption that meaning is already available in the corpora they work with and merely needs to be collected. Even DSHM and HDM, which do not use natural language, but attempt to store knowledge using direct associations between concepts, do so without the model itself offering any structuring to the input. Thus, DSHM and HDM models at the moment are better understood as association storage models rather than meaning as we discussed it. The significant advantage of VSA is that it provides us with an architecture which allows for a unified representation of knowledge. What it needs are relational forms which structure the knowledge in this architecture and make it possible to draw inference due to the presence of the selected structure.

A solution could be to attempt to implement the relational forms of the formal semantics like Montague Grammar (MG) into VSA, but given my goal, there are a few issues with that. We discussed earlier what is referred to as the *meaning barrier*—the gap between our knowledge stored in form of high-level representations and the low-level perceptions from which the knowledge is acquired. The work in this thesis is part of the larger project to fill this gap and model meaning as humans experience it—from perception to proposition. In traditional approaches of logical semantics, concept formation is linked with the theory of definition in the form of subordinate-superordinate set-theoretic relations, and it is assumed that particular characteristics which define a concept stand as already defined or designated. In other words,

"the semantics for classical predicate logic is given by models consisting of a domain of objects over which relations are interpreted; thus, the objects are assumed to be given"(Achourioti & van Lambalgen, 2011). Given our goal, such an assumption of pre-existing objects cannot be made while choosing organizing schema in VSA.

As I discuss in section 4.5.3 (pg. 53) the choice of VSA as the modeling architecture of my work makes sense exactly because VSA models are psychological first. The models like VSA "aim at modeling 'concepts in the head' rather than 'things in the world', and thus clash strongly with the ostensive anti-psychologism of MG [Montague Grammar]" (Kornai & Kracht, 2015). Even if deploying logical forms of the kind in traditional formal semantics was to be successful, it will merely result in amodal, abstractive concepts which are defined merely as sets of predicates and the model will not have a way to be expanded to allow it to bridge the meaning barrier. A similar issue will occur when we compare Kant's logic system to classical logic in section 6.3.1 (pg 85).

Hence, VSAs lacks primitives in form of the logical structure of a kind which fits in with its psychological philosophy. Such a structure needs to be designed keeping in mind not the reference to objects in the world, but to the objects, as constructed cognitively. It is for such a system that I will turn to Kant whose theory has informed much of my formulation of a proper theory of meaning and critique of existent semantic models. As we have discussed earlier, for Kant, the meaning is not plucked out of the world as many empiricists would like to believe, but rather manufactured via the forms of understanding and perception. These forms form the productive grounds and are responsible for the recursive synthesis of the manifold data from our interaction with the world that finally leads to the formation of concepts. These are understood

best not as set-theoretic logical relations between concepts—not a unity by abstraction, but unity through rules which provide lawful structuring of the manifold of our experiences. Concepts as a set of rules that then legislate the way we see the world. Concepts as prospective and not abstractive. Chapter 6: discusses Kantian epistemology and Kant's system of logic which is later implemented in my model.

4.5.2 Lack of Clarity and Methodologies

A critical issue that prevents the development of VSAs is a lack of proper methodology in the literature which allows for the systematic development and assessment of cognitive models. Consequently, evaluation is currently the weakest aspect of the research (Baroni, 2013). The reason for this is twofold. Firstly, due to mathematically obscure nature of these models, it is hard to intuitively map the many possible configurations of architecture to cognitive correlates of the mental processing. Secondly, most VSA models are created for automated processing of natural language data. Hence, during the development of these models, the important constraints were not psychological, but rather computational. For example, methodologies explored in these models are done only to conserve computational power and memory, accurate clean recall of information, etc., rather than accurate modeling of human cognition. Where psychological constraints are considered, they are done so only in regard to linguistic behavior in humans, and not general cognition.

As a result, while there do exist many differentiated models of VSA, the selection of the parameters available in VSA architecture are made in an ad-hoc manner to achieve desired computationally efficient output through 'trial and error' method without a systematic

methodology to inform the design and compare different models. This has lead to a situation where far too many parametric variables exist in the VSA modeling literature without any organized mapping to cognitive correlates that can inform the development of cognitive models in the architecture. For example, just combining of representations in VSA can be done by addition, convolution, vector multiplication, tensor product, weighted addition, vector dilation, permutation, etc. Similarly, a model can be varied based on the size of n-grams used to store relationships, whether vectors are normalized or not, or whether relationships can be repeatedly stored, the nature of input, etc. What we have as a result is a kitchen with not just too many cooks (which is good for the research in the field) but also too many tricks and no common frame of reference. Kelly & West, (2012) takes a step towards developing such a methodology by clarifying which parts of VSA are symbolic in nature and which are sub-symbolic. Extending this project requires exploration of the difference in behavior of models when these parts are changed. To achieve this, I created a model of VSA similar to DSHM in a programming language called R. The model allows for exploration of many symbolic and sub-symbolic parameters of VSA through simulations. In the next chapter, I discuss the formulation of the methodology I develop to classify and test VSA models and describe the results from the simulations I conducted to find what sort of behavior emerges when the some of the many parameters are changed.

4.5.3 Advantages of VSA

While the lack of inferential power is problematic, the statistical approach of VSA seems to be a more promising choice because it is the only approach which fulfills the more fundamental of the features of human cognition—graded meaning and neural plausibility. VSA-based DSHM

has been used to model a number of human memory effects (Rutledge-Taylor et al., 2014) including the fan effect (a delay in the recall of things with a larger number of irrelevant associations), the problem size effect in math cognition, dynamic game playing (detecting sequential dependencies from memories of past moves), and time delay learning. Inferential operations are a particularly high cognitive feature and they must supervene on these other more fundamental features. Logical models are useful to discover the precise computational operations needed to be undertaken to model a cognitive phenomenon. But their implementation must be constrained by markers of psychological behavior. Many of these markers arise naturally out of VSA-like architectures. These models thus allow for bridging the gaps that exist between competence models and performance models.

Secondly, VSA's can be used to implement different kinds of organization structures allowing for the creation of modality-specific knowledge. For example, VSA model developed by Kesorn & Poslad (2012) stores low-level visual structures within images. In fact, their model even uses a priori encoded representations to act as a specialized ontology to provide a rule-based system which guides the process of information encoding from raw images. In addition to this, VSA shows an ability to be stacked to form a hierarchy of interacting systems which can use the output of lower systems as primitives for the higher system (Kachergis et al., 2011; Kelly, 2016). This ability to store and process non-linguistic data as well, and the ability to create an interface with other VSA modules makes VSA unique in that it fulfills Jackendoff's call for the general architecture through which multiple kinds of cognitive processes can occur. They are thus also

promising to bridge the perception-conceptualization gap. Given all these features, it makes sense to stick to VSA in our attempt to design a more comprehensive model of meaning.

4.6 Chapter Conclusion

VSA seems to be a promising solution to the issue of the lack of general architecture which allows for the satisfaction of the multiple constraints that must be satisfied by cognitive models of meaning. VSA architecture is geometric in nature and allows for a representation of knowledge which is global by default. At the same time, it allows for instantiation of symbols which can be recursively bound. The models of VSA are criticized for their lack of organizational a priori structures leading to bag-of-words/concepts issue. This is a problem in the sense of both knowledge being unorganized (merely bagged) as well as it being amodal (words) without any interfaces with the world with which grounds the knowledge. While my larger project is to resolve both these issues, the work in this thesis is concerned with the former issue. Nevertheless, some of the research on VSA shows that it has a potential to allow for solutions regarding the second issue as well. A significant advantage of VSA is their natural satisfaction of the psycholinguistic, psychological and neural constraints.

Chapter 5: Simulation-based development of methodology

5.0 Chapter Overview

The chapter concerns with the methodological and conceptual clarification of VSA models. First, there exist too many tricks and no common frame of reference in the VSA research which has led to a lack of systematic methodology to assess these models. To address this, I discuss a hierarchy which relates different models with respect to the behavior of VSA that falls out of its architecture, sub-symbolic parameters and choice of symbolic encoding. The second issue is that being geometric in nature, the mathematics of the architecture obscures its relation to cognitive processes it is supposed to model. In order to bring some clarity to these aspects, I created a model of VSA similar to DSHM in a programming language called R to investigate the many parameters of VSA through simulations. The results are also presented in this chapter.

5.1 Developing a Hierarchy of VSA-based Models

The differences between any two models of VSA can be understood as arising out of differences which are either symbolic or sub-symbolic. Symbolic actions include the decisions made around how to structure, manage, store, and retrieve symbols while the sub-symbolic level, the modeler needs to decide how to instantiate symbols as vectors and symbol-manipulation as vector algebra. Because different tricks can be used to realize the same symbolic action, “the choice of the sub-symbolic feature is considered largely independent of the decisions to be made at the symbolic level” (Kelly & West, 2012). That said, it is not always clear what the impact of certain sub-symbolic choices will be over time. For example, both HDM and DSHM use vectors which as information is stored in them, grow to become heavy and difficult to move. Repeated

normalization which is a sub-symbolic feature that reduces the magnitude of vectors while conserving the angular relationships can solve this issue, but it is unclear what kind of impact it would have on the cognitive behavior of memory. Similarly, it is usually understood that the higher the dimensional size of the vector space, more is its capacity as it can store more data before it saturates. But as my subsequent simulations show, the impact of dimensionality can actually be felt in the storage of the very first bit of data. The noise in lower dimensional spaces does not come from saturation but from the fact that the orthogonality of any two random vectors in a dimensional space is inversely proportional to its dimensional size. Similarly, it is not very well understood, what the impact of a certain symbolic manipulation is on the behavior of the system, e.g., how repeated storage of the same information impacts the learning of the system. There is thus a need for a more granular investigation into the different symbolic and sub-symbolic parameters of VSA which needs to be described in the language of cognitive phenomena so as to create a framework allowing for cross-modal comparison.

Lack of methodology can also lead to inaccurate testing of VSA models. The Fan Effect—the cognitive phenomena of delay in the recall of things with a larger number of irrelevant associations—is often used to test the models of memory. Many different VSA models claim the ability to show the fan-effect to be a unique strength. But once the sub-symbolic aspects of vector space are understood properly, it becomes clear that all VSA models will be able to demonstrate the fan effect. It is thus not a good test to compare different models of VSA since they all will show this effect. Since there are many possible ways in order to encode information in VSA, what

is needed is a class hierarchy for testing models. This requires a clear demarcation of effects which arise out of:

i. the nature of high dimensional vector space
ii. specific choice of sub-symbolic operations like permutation, normalization, etc.
iii. choice of symbolic operations such as binding order, storage, etc.

Different models of VSA thus need a feature-based hierarchy to inform the methodology for comparing these models. I propose to classify models across the three sources of changes mentioned above (Figure 6).

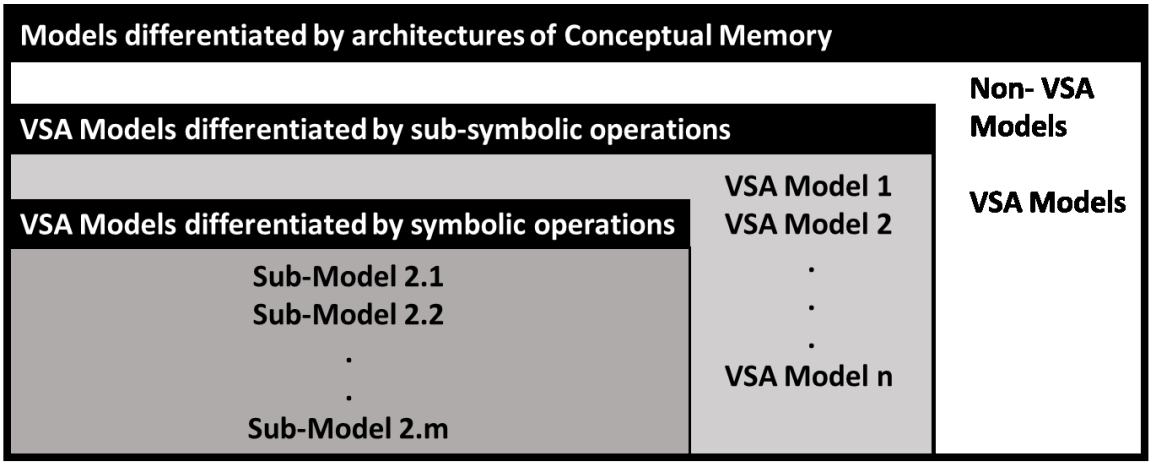

Figure 6: Methodological Clarification for testing Memory models

In this hierarchy, all sub-models lower in the hierarchy will pass the tests which are used to differentiate their parent model (Table 2). With this in mind, I proceed to examine some of the features of VSA at different levels of analysis.

Table 2: Test Behavior

	Test 1 (e.g. Fan Effect)	Test 2 (e.g. Learning Rate)	Test 3 (e.g. Recall Error)
Non-VSA Model	X	-	-
VSA Model	✓	-	-
VSA Model 1	✓	X	-
VSA Model 2	✓	✓	-
Sub-Model 2.1	✓	✓	X
Sub-Model 2.2	✓	✓	✓

5.2 Movement of Vectors and Fan-effect (Architecture Level)

In Vector Space, storage of every local relationship changes the relationship of the involved vector globally, and thus results in a realistic model of knowledge (Franklin & Mewhort, 2015). A way to understand this would be to imagine randomly placed balls on a table- **A, B, C, D**. If we move B towards A, we don't just change the distance between **A** and **B**, but also distances **B-D** and **B-C**. If we also move **C** towards **A**, not only will the distance between **A** and **C** get reduced, but also between **C** and **B**. The movement will also possibly change the distance between **D**, **B,** and **C**. (Figure 7)

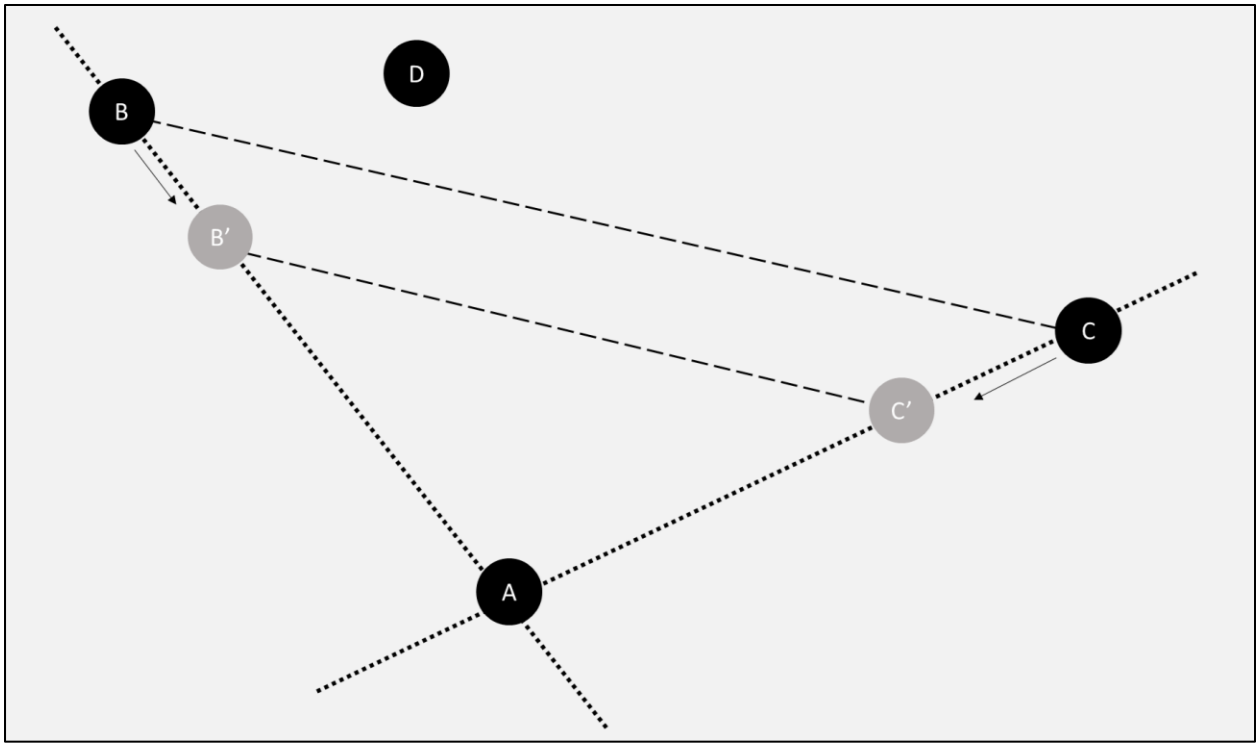

Figure 7: A Balls and Tables mode of Vector Space Encoding

To demonstrate this behavior in vector space (dimensional-size of 1024), I ran a simulation with sparse data on the model of DSHM that I created in R language. Sparse data simulations aim to highlight interesting behavior in the system without use of large datasets. R is a programming language and free software environment for statistical computing which has recently gained much popularity.

I looked at how the angle between two target concepts—*car* and *truck*—changes as I add in the system their relationship with other concepts—*wheel*, *drive*, *steer* and *road*. Vectors were initialized for each of the concepts and the four concepts are added through vector addition (superposition) to the vectors of both of target concepts sequentially. The initial angle is ~90° between all vectors indicating a cosine of 0, i.e., no similarity.

Table **3** shows the angles between the vectors after each iteration.

Table 3: Variation of vector angle for target vectors

Concept Pairs	Initial Angle	Angle after addition of respective concepts			
		Wheel	Drive	Road	Steer
Car-Wheel	88°	44°	52°	61°	60°
Car-Drive	93°	93°	57°	59°	65°
Car-Road	93°	93°	95°	63°	69°
Car-Steer	85°	85°	83°	87°	60°
Truck-Wheel	92°	47°	55°	64°	62°
Truck-Drive	90°	90°	55°	58°	64°
Truck-Road	91°	91°	94°	62°	68°
Truck-Steer	89°	89°	85°	89°	62°
Car-Truck	**90°**	**61°**	**50°**	**44°**	**39°**

Note that at no point was the vector for *Car* added to that of *Truck*, or vice versa. Yet, as we add the vectors of *wheel, drive, steer* and *road* to the vectors of the target concepts *Car* and *Truck,* the angle between them keeps getting reduced from initial 90° to ultimately 39°. So, storing information about cars and trucks will in turn automatically impact how cars and truck relate to each other even when no information relating the two is explicitly stored. Thus, in VSA, every time some information is stored, its impact is transferred globally.

But this kind of movement, which has a global effect, leads to the fan-effect to become not a special feature of some models of VSA, but all the models created under the paradigm. The *fan effect task* (Anderson, 1974) is a recognition memory task. In the original fan effect task (Anderson, 1974), each sentence is of the form “the person is in the location” where the person

and location vary from sentence to sentence (e.g., “the hippy is in the park”). The fan of a concept is the number of different concepts related to that concept. For example, if “*hippy park*” is the only relation in a study set that mentions *hippy*, then *hippy* has a fan of one. If participants learn that there are four people in the *park* during the study phase, then *park* has a fan of four. The fan effect refers to the finding that participants are slower to recognize or reject sentences that contain concepts that have a higher fan. The fan effect illustrates a fundamental principle of human memory: the availability of a piece of information in memory with respect to a cue is a function of the probability of that piece of information conditional on the cue. If the participants learn four facts about the concept and then they are given the concept as a cue, each of those facts have only one chance in four of being the relevant fact to retrieve. The retrieval time from memory will reflect that one-in-four probability. In case of VSA models, the similarity between vectors is taken to contribute to the reaction time.

In the Car-Truck example, we see that as the fan of the concepts of *Truck* and *Car* increases from one to four through subsequent addition of concepts—*wheel, drive, road* and *steer,* the mean angle between the concept and the fan elements increases as well which would result in an increasing reaction time during recall. (Figure 8). The same is visualized for a fan of two and three in Figure 9.

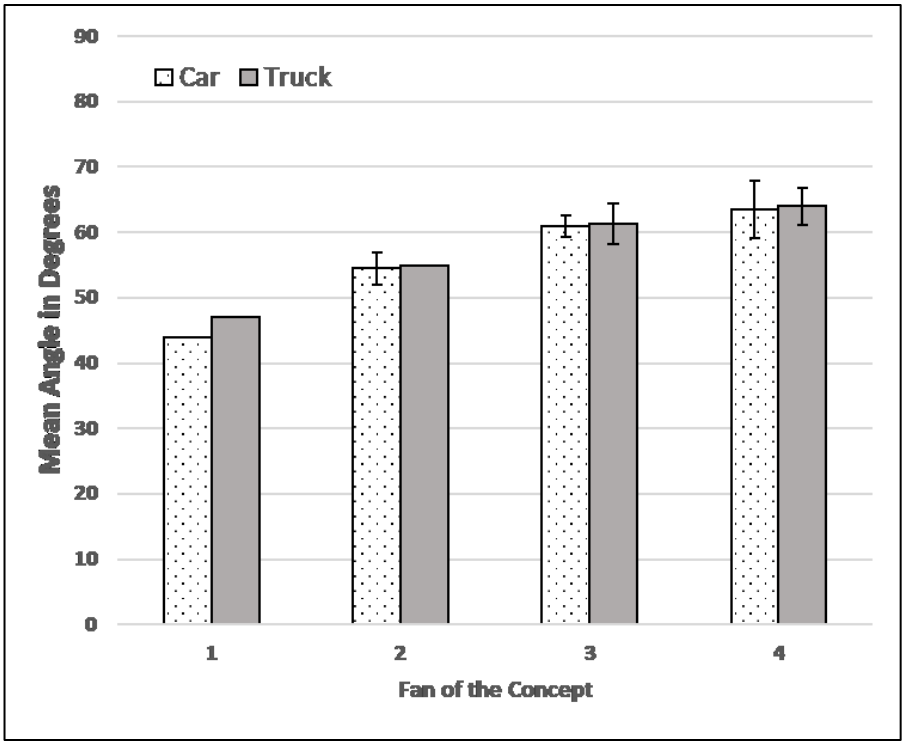

Figure 8: Fan effect

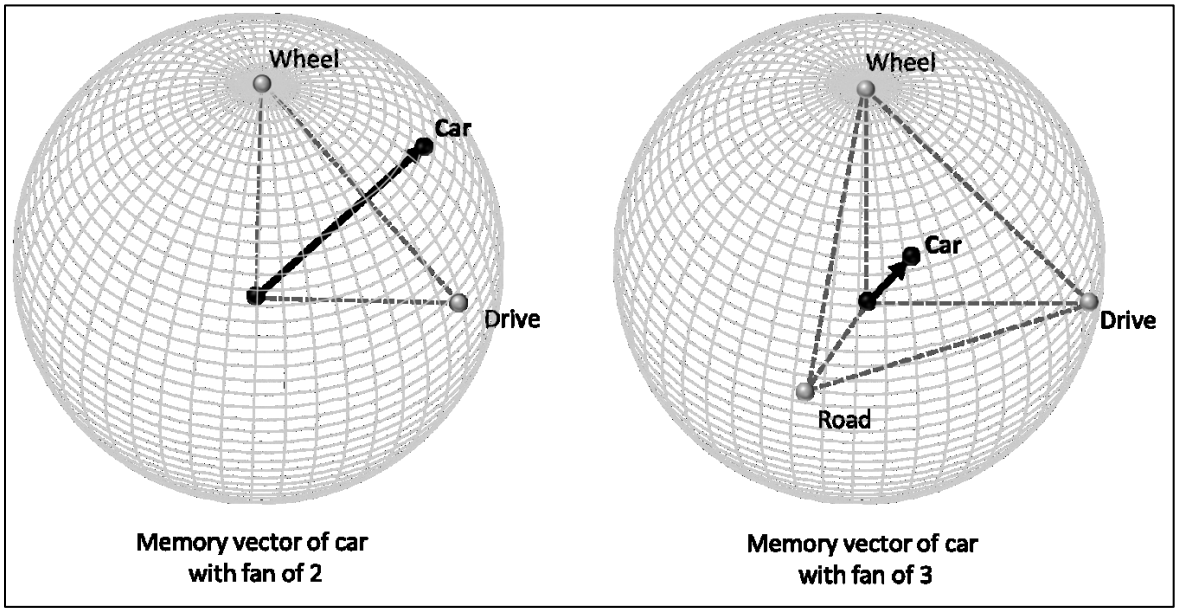

Figure 9: Visualizing angles for concepts with a fan of 2, 3

Since this effect is not specially encoded in the model but arises naturally out of the architecture, it is common across all VSA models. And thus, while the test for the existence of Fan effect is useful to compare VSA models to non-VSA models, it is not a meaningful test to differentiate between different VSA models.

5.3 Orthogonality and Dimensions (Sub-Symbolic)

As discussed earlier, VSAs are implemented in high dimensional space. But this generates a question of how many dimensions are sufficient to model a phenomenon? Most modelers agree that as we add more data, the lower dimensional systems become saturated and noisy faster than the high dimensional ones. What is less understood is why this happens, and how can we quantify the noise. For this reason, it is important to understand the effect of dimensionality on the models.

5.3.1 Approximate Orthogonality

High dimensional space is preferred for VSA because of a very interesting property of high-dimensional vector space—any two randomly generated vectors in hyperspace are most likely to be approximately orthogonal, i.e., the angle between any two randomly selected vectors in hyperspace is most likely to be close to 90 degrees. Cosine of angles between vectors can vary from one (for an angle of zero) to zero (for the angle of 90 degrees). This is important because it allows for the angle between vectors to signify the similarity between the concepts they represent. Hence, the creation of a new concept automatically leads to a vector which is orthogonal (hence dissimilar) to all other vectors, and only after some information is stored in the system about that concept is its corresponding vector non-orthogonal with other vectors. To

visualize the effect of dimensions on orthogonality between vectors, I have simulated a random generation of a 100 vectors pairs for vector spaces which differ in their dimensional sizes. The dimensions increase exponentially in each step (from 2^1 to 2^{15} i.e., from 2 to 32,768). While the average angle between the 100 pairs is around 90 degrees for all dimensional sizes; for the higher dimensions, the spread of the angles across the pairs reduces significantly. For just 2 dimensions, the angles vary from 0 to 180 degrees, but for dimension size of 256 (or 2^8) onwards, the angles do not go beyond the interval of 80 to 100 degrees (Figure 10). We observe two patterns:

i. newly initiated vectors are not entirely but approximately orthogonal, and hence have some similarity between them by default. This leads to noise in the data storage from the very beginning.
ii. As the dimensionality is increased, this initial noise decreases because high dimension spaces have a higher number of possible vectors which are approximately orthogonal.

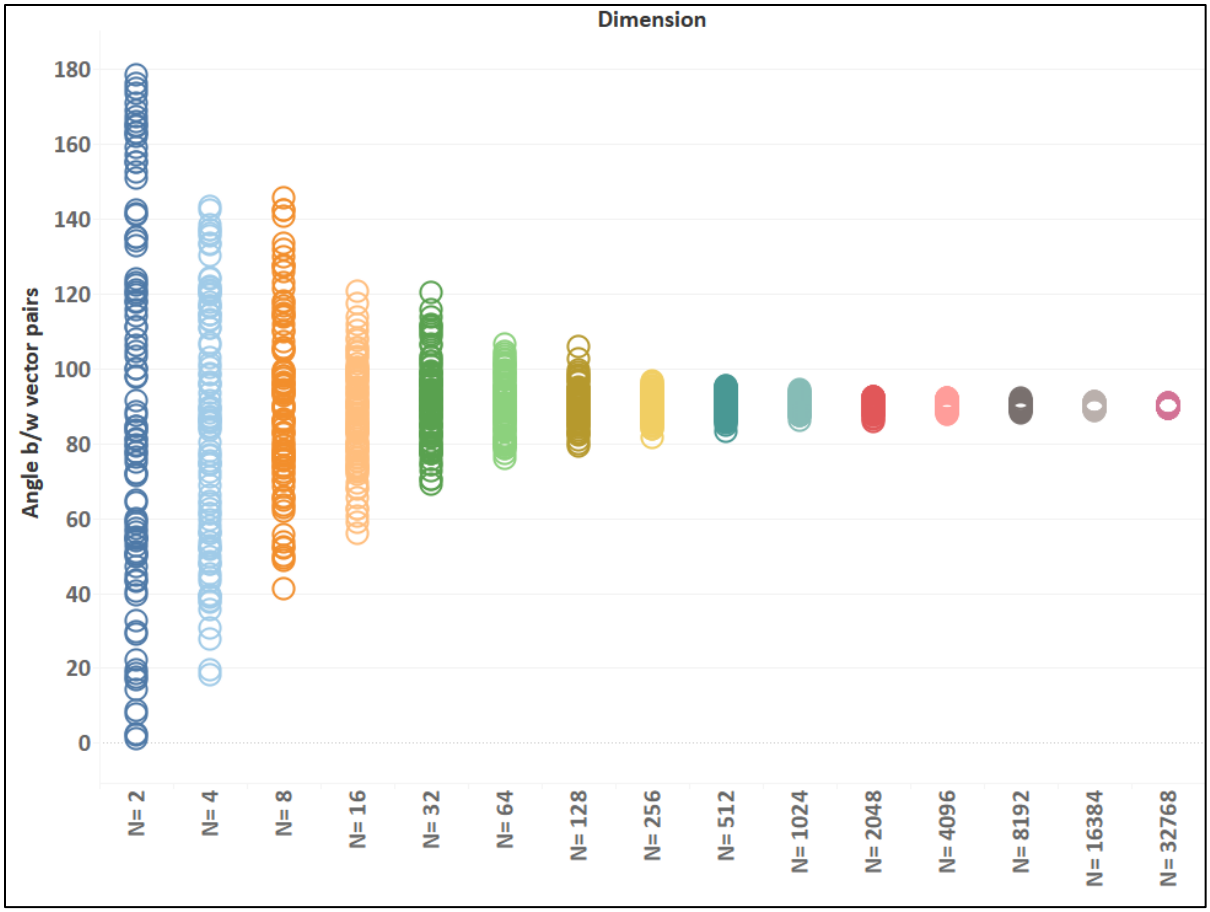

Figure 10: Randomly generated vector pairs in hyperspace are most likely to be approximately orthogonal

5.3.2 Reliability of Vector Spaces

Figure 11 shows the average similarity between the vector pairs and error margins for the averages computed. For the angle between two vectors to indicate significant similarity, it must be beyond the error margins indicated by the average similarity and standard error for those averages. For example, at the dimension of 512, where we see the 100 random pairs vary only from 80 to 100 degrees and see an average similarity of ~5% with 5% error margin. Both of these are a function of the number of dimensions.

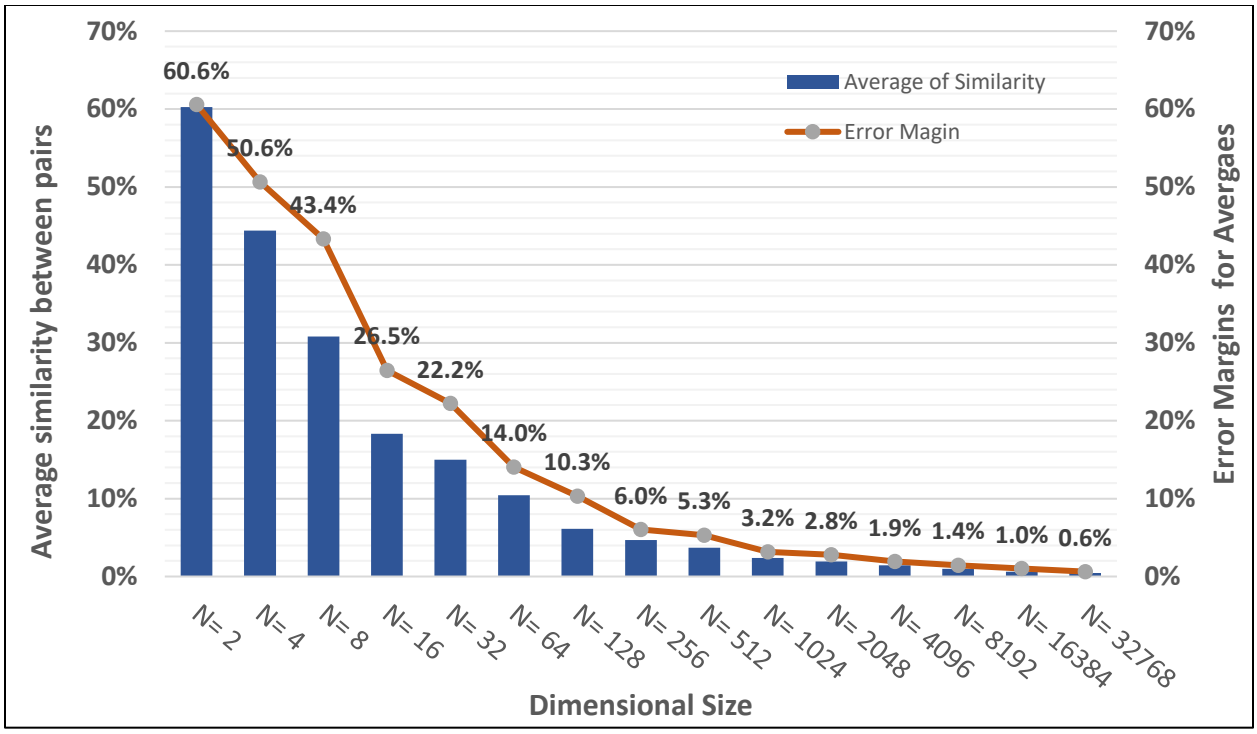

Figure 11: Average Similarity Across Dimensions

This method thus gives us a way to decide dimensionality based on what the error margins we are comfortable accommodating in our system. It also allows us to judge whether the differences in similarity that we obtain in a model are due to actual data or can be due to initial noise. Secondly, it shows that the noise in the vector space is not due to the excessive data, but due to the dimensional parameters of vector space that get fixed right at the beginning.

5.3.3 Robustness of VSA

The result of such properties of hyperspace is that any significant similarity (angle of less than 90 degrees) between two vectors is almost always an indicator that at some point one of the vectors was added to the other. In fact, it takes over 100 additions of random vectors (of dimensionality 500) to a vector before a non-constituent vector is accidentally recognized as a

constituent (Widdows & Cohen, 2014). As a result, in high dimensions, new elements can be created automatically with almost no effort, and almost no danger that they will clash with pre-existing elements. This is crucial to the robustness of VSAs, because it means that we almost never encounter accidental similarities between elements that have nothing in common.

5.4 Normalization (Symbolic Level)

A sub-symbolic parameter for VSA models is whether or not vectors are normalized or not. Initially, the memory vector of a concept in VSA has an amplitude of 1 unit. As more relations are added to the memory vector, the amplitude increases, and the vector becomes heavier. What this means is that the concepts show greater movement in the beginning and as more information is added, their movement becomes more difficult. This can be avoided if vectors are normalized after every addition. Normalization is an operation which conserves the angular distance of the vectors from other vectors while reducing the amplitude of the vector back to 1. It is thus meaning (reflected in the angles between the vectors) conserving.

To demonstrate this, I simulated 100 additions to a vector (N=1024) under two conditions. In the first condition, the resultant vector was normalized after every addition, while in the second condition, the vector was not normalized. Figure 12 shows the angles by which a vector varies due to an addition of information across multiple additions. For unnormalized vectors, we can see that the angular variation decreases as more data are added. In case of a normalized vector, the angular variation is steady at around 45 degrees.

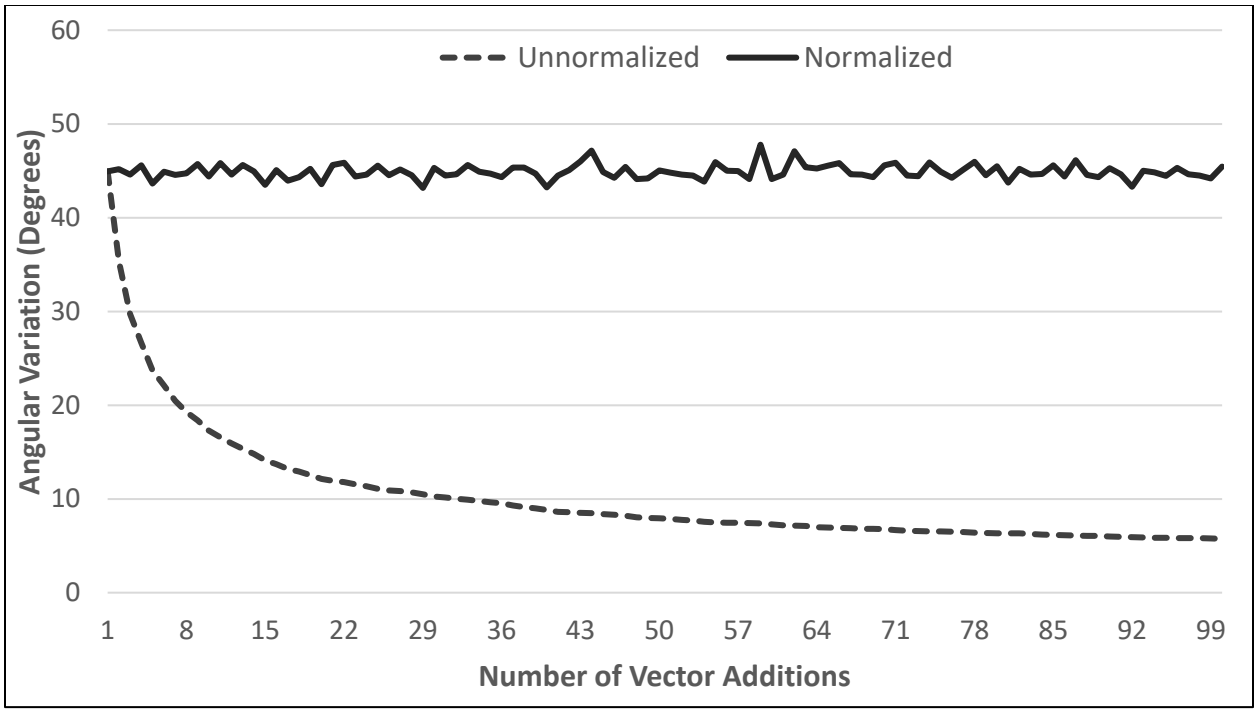

Figure 12: Angular variation of Memory Vector on the addition of data

Thus, whether the vectors are normalized or not corresponds to a very different learning rate over time. One can even control the rate of normalization such that the rate of amplitude gain is reduced but not altogether eliminated. Variation of the normalization parameter can thus be used to model different kinds of memory. It should be expected from the models of long-term memory to show some kind of eventual slow-down in the movement of knowledge which has been built over a long time and thus corresponds to the unnormalized VSA storage. On the other hand, working memory seems much lighter and able to act fast with new information about already-existent concepts and thus corresponds to the normalized memory system. This parameter is added to my final model (named Kantian-HDM) and allows it to simulate and test different kinds of memory systems. I discuss the model in the final chapter of the thesis.

Another interesting thing that occurs due to variation in normalization parameter is how fan effect is manifested in these models. As discussed above, all VSA models show fan effect. When the normalization parameter is turned on, what does change is the angle between the arms of the fan. For unnormalized models, the angle between the arms is approximately equal (Figure 13), but for the normalizing models, the angle in lesser (higher similarity) for the more recently added arm and increases for the earlier added vectors (Figure 14). This behavior is significant in that it makes a testable prediction regarding human memory—if normalization corresponds to how the brain stores information, we should have not just the fan effect, but a recency effect with regards to its different arms as well.

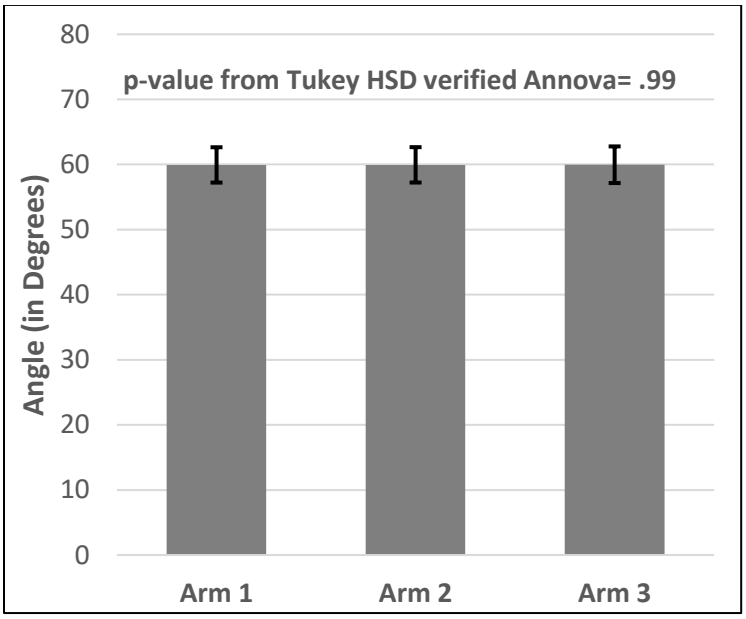

Figure 13: Fan= 3, Mean angles between arm and concept (Normalization=FALSE)

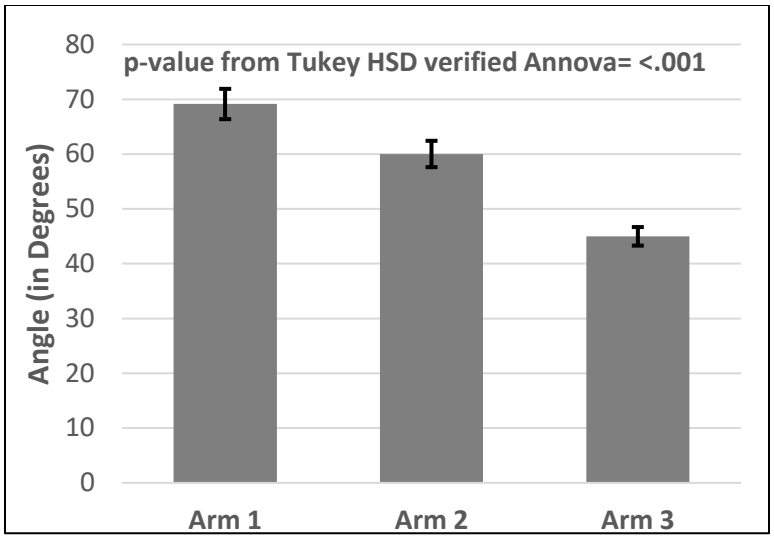

Figure 14: Fan= 3, Mean angles between arm and concept (Normalization=TRUE)

5.5 Chapter Conclusion

The chapter presented a new way to classify the models of memory with respect to VSA in the form of a hierarchy based on its architectural, sub-symbolic and symbolic aspects. Furthermore, investigation through simulations of models which vary parameters at these different levels sheds much-needed light on how VSA models are able to demonstrate the behavior that they do. The discussion allows for not just better assessment of VSA models but also better design.

Chapter 6: Kantian Epistemology

6.0 Chapter Overview

The models of VSA are criticized for their lack of organizational a priori structures leading to bag-of-words/concepts issue. Part of this problem is in knowledge being unorganized (merely bagged) in these models. The chapter introduces Kant's epistemology with an aim to motivate its use for resolution of issues in VSA models of meaning. Kant approached the issue of conceptual knowledge from perception instead of language. Convinced of the necessity of explication of how both content and concept arise, Kant proposes two separate but linked faculties of mind which provide first the spatial and temporal forms to the material from the senses, and then categorial logical forms that ultimately leads to constitution of the object concept. Kant calls this body of logical forms *transcendental logic*. It is this logical framework that fits the goals of this thesis and the psychological and geometric underpinnings of VSA architecture. I end the chapter with an assessment of Kantian categories.

6.1 Situating Kant

As we discussed earlier in section 2.2.1 (pg. 19), a fundamental distinction that has arisen from research on meaning is that of the psychology of language vs the psychology through language. A similar issue is present in epistemology. The reality of the structure of the world and the structure of our mind are combined in a way that makes it very difficult to separate the two. Thus, it is important to remember the distinctions between:

i. the structure of reality
ii. the structure of the human mind

iii. human knowledge/experience created in the interaction of the above two
iv. the specific structure of human linguistic abilities
v. the linguistics expressions which express our knowledge under the linguistic constraints

Aristotle's leap was from (v) to (i) when he tried to derive ontological truths, i.e., truths about how reality really is, through an analysis of language in which he believed it was reflected. Most VSA-like models of meaning that we discussed above, especially the ones which explicitly use natural language to model meaning make the leap from (v) to (iii). Kant, on the other hand, is more difficult to pinpoint. He puts the knowledge of reality—things in themselves (i), out of human reach. For him, what we can know are truths about appearances that are born out of the interaction of reality with the mind (iii). And thus, for him, the question was what we can say about the structure of human mind (ii) given that all we have is an experience, and thus only an appearance of reality. In doing so, he set forth his enterprise to be that of describing the necessary architecture of the mind, and from there, the necessary limits of its ability to use reason to produce knowledge of things it does not have a direct access to. Kant's leap can thus be seen as from (iii) to (ii). A model of a complete theory of meaning requires an explanation how we obtain (iii), that is, knowledge of the structure of the world (i) via a theory of the structure of our mind (ii). It is this structure which constitutes the conditions of having meaning. Kant's theory thus makes a promising candidate for my enterprise here.

Strawson, who puts Kant, along with Descartes, Leibniz, and Berkeley, in the category of "descriptive metaphysician", expresses their goal as "to lay bare the most general features of our conceptual structure"(Strawson, 1959, p. 9). Kant virtually never alludes to the language faculty

in his critical work; for him, it is in our knowledge structure that we can find the key to the transcendental self. But he nevertheless works abundantly with the propositional features of language such as subject-predicate relationships. Nevertheless, he did not restrict himself to merely conceptual analysis but also to its connections with perception. And in doing so his work went right into the issue of *meaning barrier* I discussed earlier — the gap between our knowledge stored in form of high-level representations and the low-level perceptions from which the knowledge is acquired.

Kant claimed that these properties of mind lay innate in even the lowliest intelligent of all humans**.** Innate structure, which is a core feature of Kantian philosophy, is more recently pushed by Chomsky's brand of linguistics. In fact, Williams, (1993), in his Book titled *Kant's Philosophy of Language: Chomskyan Linguistics and its Kantian Roots,* explicitly talks about the connection between Kant's account and Chomsky's theory. In reference to *the Critique of Pure Reason*, stating that "the precursor of the more specialized positions of both Humboldt and Chomsky may be elicited from the pages of Kant's masterpiece." That said, crucial differences exist between the two. Chomsky locates the generativity of human cognition in a recursive syntactic feature (Merge) which evolved in its entirety through a mutation, and which is responsible for the uniqueness of human cognition. This single mutation account is one of the biggest criticisms of the *Merge* narrative. On the other hand, Kant's innate structure consists of an elaborate list of formal functions which organize our knowledge and give it the nature it has. These functions are the recursively generative operations which construct meaningful representations which we come to refer to as concepts. They are thus not syntactic in nature. Because there are multiple

kinds of these unifying functions, each of which add a certain kind of richness and structure to our knowledge, it allows for a possibility of a gradient evolution of the multiple kinds mental powers which come together to make humans unique.

According to Strawson, "no philosopher understands his predecessors until he has re-thought their thought in his own contemporary terms; and it is characteristic of the very greatest philosophers, such as Kant and Aristotle, that they, more than any others, repay this effort of re-thinking.". Cognitive science has progressed in a similar way where each advancement in technology, from the abacus to gears to computers, has been used as a metaphor to re-conceptualize mind in an informative manner. As we make advances in new ways of representing knowledge, it is essential to understand how the historical work in philosophy of epistemology fits in and can be understood through these new methods, and might help us bind these methods into a more holistic model of meaning. In this thesis, I use the features of vector spaces to reconceptualize Kant's theory.

6.2 The Critique of Pure Reason

Immanuel Kant has been called one of the most influential philosophers of all time. His famous work, *The Critique of Pure Reason* (CPR) (Kant, 1781), is an inquiry motivated by the question of what it is to be a knowing subject, and consequently, what can be known and how.

His remarkable insight was to reverse the empiricist assumption that our knowledge of the world conformed to its object, and instead claimed that it is the object, as we perceive it, that must conform to our most basic conceptual structure. For him, this was equivalent to the Copernican revolution—the reversal of the assumption that the earth was at the center of the

cosmos. This was no small move and resolved some tension between the rationalist and empiricist traditions. He envisaged a middle ground of "appearances" between reality and complete illusion; a middle which is real in that it is tethered to external reality and obtains its content from it but at the same time, is structured through the innate tools of mind and formed through it.

For Kant, we can only have metaphysics confined to the ontology of mind as it must be to have the experience of the world as we do. As Kant states, "the proud name of an Ontology that presumptuously claims to supply, in systematic doctrinal form, synthetic a priori knowledge of things in general…must, therefore, give place to the modest title of a mere Analytic of pure understanding."[3]

This is what makes Kant's approach especially interesting. Centuries before linguists used a constraint-based approach to theorize on rules of language, Kant had begun his critical analysis of human cognition by deriving constraints from the insights into the nature of human experience. While the structure he proposes might not be exhaustive, nor is his notoriously difficult to read text always clear on how this structure is realized in mind, his analysis opens directions for development that have been repeatedly recognized to be both novel and fertile.

6.2.1 *A Priori* Faculties, or Innateness

[3] Critique of Pure Reason, A 247

Taking a leaf from rationalist traditions, Kant uses the concept of *a priori*; and expands it to mean more than just pre-experiential. For him, *a priori* functions provide the very form to our knowledge, and hence there cannot be any knowledge that does not have these forms. As a result, these innate forms must have a quality of *universality* and *necessity.* For anything to qualify as a primitive, it must be both universal and necessary. We thus see a strong criterion for selection of primitive forms. *A priori* forms are the Kantian answer to the question we raised earlier regarding meaningfulness—what are the conditions that make meaning possible? Empirical knowledge is always contingent, susceptible to change and hence neither universal nor necessary and thus separable from the *a priori.* He located these *a priori* forms in the form of sensibility and understanding.

For Kant, the sense of the object is literally the first sensory interaction with the world, and sensibility allows us to make this first contact. The forms of understanding, on the other hand, create the conditions through which the *sense* manifold is *referred* to the concepts in the mind. Both these processes were equally important for Kant and his belief is distilled in his famous quote "Thoughts without content are empty, intuitions without concepts are blind." (Kant, 1781). Convinced of the necessity of explication of how both content and concept arise, Kant proposes two separate but linked faculties of mind:

i. Intuition: a passive faculty responding to the world and sensing it. It is the source of the a priori forms of space and time present in our perception.
ii. Understanding: an active faculty which cognizes the world by referring the sensations to concepts using a priori logical forms.

Under what he calls the Aesthetics, or study of sensibility (the power of intuition), he observed that all our experience is necessarily structured in terms of space and time and identified these as *the forms of Intuition*. Space and time are thus necessary and universal forms of sensible experiences. The first moment of representation consists of combining the manifold of appearances through *a priori,* intuitive, unity which is characterized by spatial and temporal forms. Space and time, as pure intuitions, stand as the "first ciphers for the appearances that supply the principles of their situational coordination as coexisting and successive" (Kant (1781), 1995). Such an encoding for Kant is the first step in making the still yet indeterminate given into a determinable appearance. According to Kant, we have little or no consciousness of this stage. One can understand this as a mode of experience where the object is yet to be defined with respect to the subject. One might get close to imagining such a state if one think of the vague perception one has of the surrounding when distracted by a deep thought. It is not that there is no perception, but it is of the kind where one does not have any salient objects, but just a rhapsody of appearance. The manifolds, now structured in spatial and temporal forms is yet to be subsumed under the concepts, and through it, made into a meaningful experience regarding objects of perception.

It is under what he calls *Logic*, the science of Understanding, that Kant discusses the next steps which finally result in objective knowledge. Objective for Kant is for cognition to have an object. The objective validity, that is valid with respect to our previous knowledge of the object we are perceiving, results from invariable pure concepts of understanding known as *the categories*. As mentioned earlier, the manifold of perception, given structure through spatial and

temporal form only provide material which is yet to become knowledge. At this moment, Kant makes his next big move by claiming that knowledge is made possible not merely through these empirical data, but through the act of judgment in accordance with *a priori categories*. And the key insight is that one and the same function, that of judgment under categories, unites both the manifold of sensory intuition into a subjective thought and then unites this subjective experience further with conceptual knowledge. The unification can be either subjective—i.e., if the combination is relative only to a single subject in a specific situation; or it can be objective—i.e., if the combination is itself an affirmation of a universally (throughout one's gathered knowledge) valid correlation between the subject and predicate that is valid independently of all particular temporal circumstances.

> "The same function which gives unity to the various representations in a judgment also gives unity to the mere synthesis of various representations in an intuition; and this unity, in its most general expression, we entitle the pure concept of the understanding. The same understanding, through the same operations by which in concepts, by means of analytical unity, it produced the logical form of a judgment, also introduces a transcendental content into its representations, by means of a synthetic unity of the manifold in intuition in general"[4]

---

[4] Critique of Pure Reason, A 79 /B 105.

The transcendental logic signifies nothing other than the variety of ways of a possible unification of the manifold of perception in consciousness. These modes of unification occur, not through sensibility, but through judgment which has multiple categories.

6.2.2 Categories

CPR contains first a general argument selecting particular judgment forms (called the Metaphysical Deduction), followed by detailed arguments aligning particular logical forms with particular cognitive functions directed toward the constitution of objects (Transcendental Deduction). This thesis is part of a larger project to create a VSA system which implements both. In this document, I focus on the former of the above and attempt to implement the logical relations in VSA.

According to Kant, it is this act of understanding, which he calls "transcendental logic", which makes it possible to acquire empirical knowledge. Kant called these logical relations *categories*. Through this move, Kant anticipates the 19th and 20th-century developments in logical theory (N. K. Smith, 1918). The *categories* are key to the organization and logical structuring of concepts. For Kant, they are innate functions of human cognition and knowledge, and makes it possible to subsume the experience of sensibility into our conceptual knowledge of the world. When we perceive, we cognize an object not merely as it is, but relative to the rest of our knowledge of the world. In this way, our perceptions get meaning, and empirical knowledge gets structured into concepts.

To derive these innate *categories*, Kant first asks what kind of judgments we can have in which any two concepts can be related with respect to each other. Kant derives a list of four

categories each of which has three kinds. All relations between concepts adhere to a type from each category (

Table 4). These *categories* are principles that organize our experience in terms of universal relationships such as quality, quantity, cause and effect and so forth, and are thus empty of any specific content.

Table 4: Table of Categories

Category	Type	Logical Form
Quantity	Universal	**All** A is B
	Particular	**Some** A is B
	Singular	**A** is B
Quality	Affirmative	A **is** B
	Negative	A is not B
	Infinite	A is **a non**-B
Relation	Categorical	A is B
	Hypothetical	If A, then B
	Disjunctive	A is B **or** C **or** D
Modality	Problematic	A may be B
	Assertoric	A **is** B
	Apodeictic	A must be B

As mentioned earlier, there are two kinds of unifications undertaken through these categories. The first kind of unification acts on the manifold of intuition, organized through the temporal and spatial forms, to determine the elements which belong together and gives us a subjective unification—combination relative only to a single subject in a specific situation. To take Kant's own example, picking up a body leads to a subjective phenomenal experience of a body being heavy. This arises out of unification of the element of the visual and haptic perception of the body in the interaction with a body. This experience is subjective and specific to the particular time and space and the source of the sensory experience. What this means is that the

judgments made at this state of cognition are merely those of subject's instances of individual perception. The data are coordinated but merely as a result of subjective rules of association and empirical concatenation, and hence established only within the specific spatial locale and temporal moment—in the *there*, *then* and *this* of the experience. This leads to a subjective experience of the body being heavy.

The second is a unification of the perception under concepts. Judgment through categories allows us to bring the subjectively valid perception into universal objective validity which results in apophantic judgment—truth-valued proposition. Through these acts of unification, thought transcends its limitation in specific spatiotemporal experience, and establishes the specific experience as part of a larger systematic whole consisting of many interrelated objects. Through *synthesis* under the logical form (in this case *is*), it translates to the knowledge that "body is heavy". Under Kantian *categories*, the relationship is quantitatively *singular*, qualitatively *affirmative*, relationally *categorical*, and modally *assertoric,* and is experienced as such. The goal of these judgments is not the isolation and establishment of the specific temporal fact, but the subsumption of the individual (the particular body) under a concept (bodies) by which the individual fact (the body is heavy) is determined as universally valid within the order of the natural world (bodies are heavy). Thus, it achieves a lawful validity, and can be understood as a result of having occurred out of a universal property of everything that can be considered a *body.* All this is achieved through the application of *a priori* categories.

Thus, contrary to classical philosophy's task of determining the properties of the objective realm of being, Kant locates his goal in the determination of those functions of cognition through

which the object of experience it itself made possible, and through which we build our constitutive structure of the natural world. As a result, objectivity, giving of the object, arises not through recognition and retention of vivid impressions, but through a synthesis made possible by a set of relations recurrently operating on various contents of presentations (Sloboda, 1995). In this sense, the object emerges as nothing other than "that something, the concept of which expresses such a necessity of synthesis."[5]

6.2.3 Faculty of Reason

Aside from Intuition and Understanding, Kant also mentions a faculty of Reason. For Kant, it is the faculty of *Reason* through which truth of judgments is determined. For Kant, senses do not judge at all, and hence cannot error. Judgments occur in the understanding and make it possible to have experience. The error of the judgments, or the truth (understood as agreement of different cognitions) of experience, is determined through the faculty of reasoning. "Whether this or that putative experience is not mere imagination [or dream or delusion, etc.] must be ascertained according to its particular determinations and through its coherence with the criteria of all actual experience"[6]. Both deductive and inductive inferences are located under this faculty.

---

[5] Critique of Pure Reason, A 106

[6] Critique of Pure Reason, B279

For Kant, while understanding unifies appearances through rules, reason unifies rules of understanding through principles. For example, while the category of cause (hypothetical) allows us to unite cause and effect, the principle that everything that occurs has a cause comes from *Reason.* Thus, reason requires categories of understanding to operate on. "Reason initially never deals with experience or any object, but deals with the understanding in order to provide the understanding's manifold cognitions with a priori unity through concepts."[7] We thus see a dependency of inference abilities over the application of categorical structure. Whereas Understanding unites a perceptual manifold into a conceptual whole, Reason unites conceptual judgments into an overall whole such that they all agree with each other. Figure 15 shows how the three faculties relate to each other.

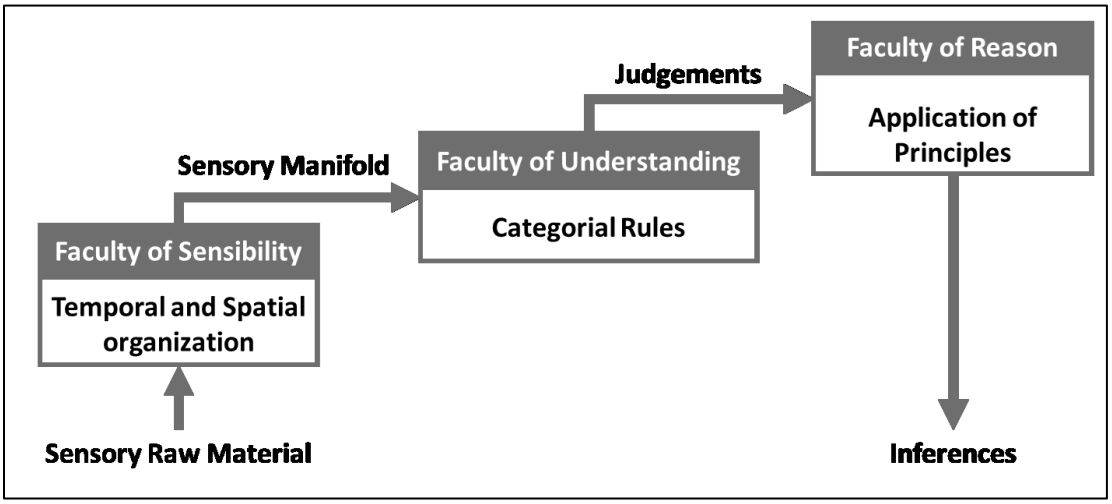

Figure 15: Kant's innate faculties

[7] Critique of Pure Reason, B359

If one were to commit to this architecture, it would be essential for the system to encode information through categorial rules before it can use the representations so created to derive inferences from it. Following the rationale, for VSA systems to develop inferential abilities would require it to first encode concepts via a categorial model and then apply principles of reason over them. The same is the goal for the model I develop.

6.3 Assessment of Kantian Categories

6.3.1 Transcendental Logic vs Traditional Logic

Kant's transcendental logic of categories does not seem to be a logic in the modern sense at all: no syntax, no semantics, no inferences. But the reason for this is because Kant's motivation for the development of the system of categories derived from the table of judgments was not the same as those behind the development of classical predicate logic. "For Kant, judgments somehow play a role in constituting objects from the sensory material, so that it seems wrong to take objects as given from the outset" (Achourioti & van Lambalgen, 2011). Claiming that "Kant's 'transcendental logic' is a logic in the strict formal sense, albeit with a semantics and a definition of validity that are vastly more complex than that of first-order logic", the authors of the 2011 paper proceed to develop a formal semantics to describe transcendental logic through a body of logic known as geometric logic. The details of the undertaking are beyond the scope of this thesis. What I will discuss though is Kant's motivation behind categories.

"Kant selected forms of judgement for inclusion in the Table not because these were sanctioned by traditional logic, but because they play a role in 'bring[ing] given cognitions to the objective unity of apperception', that is, to relate our representations to objects"(Achourioti &

van Lambalgen, 2011). It is this conceptualization of relational forms that make transcendental logic different from standard predicate logic. This is also what makes Kantian relations compatible with deeply psychological semantics. We discussed in section 4.5.1 (pg. 49) a similar issue with regard to the opposing philosophies of VSA architecture and first-order logic.

Transcendental logic deals with the construction of objects of experience and the objects in Kant theory have internal structure. The objects are not those in the world (transcendental) which can be defined by the description of sets they belong to, but objects represented internally to be experienced as appearing to exist outside. It is these complex objects which his logic attempts to lay the semantics of. The categories are the rules through which such a construction takes place via acts of judgment.

6.3.2 Criticality of Categories

A further analysis of the judgments seems to suggest that some of the categories might be more fundamental than other categories when forming relations between the concepts. Kant insisted that the same kinds of judgments through which we relate concepts are the cognitive actions through which we give the object to our experiences. But not all of them might be critical for forming conceptual relations even if they might be for the perception of the object. The *Quality* category which indicates whether A is positively or negatively associated with B seems to be necessary for any judgment about A with respect to B. Similarly, the *Relation* category also contains essential judgments without which any judgment about A and B does not make much sense. But one can conceive of judgments which contain these two relations but remain

ambiguous about whether the relation is universal or particular, necessary or merely possible (see

Table 5). As I discuss in later sections, VSA allows us to estimate the Quantity categories from the judgments based on other relations stored in the memory.

Table 5: Critically of Categories for Conceptual Relations

Category	Type	Logical Form	Sense of Judgment without the Category	Criticality of Category for concept
Quantity	Universal	**All** A is B	All Dogs are loyal, or some dogs are loyal, or a dog is loyal	Non-critical; some knowledge about dogs being loyal
	Particular	**Some** A is B		
	Singular	**A** is B		
Quality	Affirmative	A **is** B	Dog is hungry, or Dog is not hungry, or Dog is non-hungry	Critical. knowledge is not possible without the category specification
	Negative	A is not B		
	Infinite	A is **a non**-B		
Relation	Categorical	A is B	Dog is hungry, or if there is a dog, it is hungry, or dog is either hungry or not hungry	Critical. knowledge is not possible without the category specification
	Hypothetical	If A, then B		
	Disjunctive	A is B **or** C **or** D		
Modality	Problematic	A may be B	Dog may be hungry, or Dog is hungry, or Dog must be hungry	Non-critical; Some knowledge about dogs being hungry
	Assertoric	A **is** B		
	Apodeictic	A must be B		

6.3.3 Exhaustivity of the system of categories

There are some concerns regarding the exhaustivity of Kantian categories when it comes to expressing all knowledge. For example, do categories allow for object identity or constancy? The Kantian Quantitative category of the type Singular (A is B) can arguably be understood as constituting a judgment on a specific token of A (The A is B). It might be possible to instantiate

tokens using this understanding of the *Singular* category type. The concerns around whether the categories can relate objects temporally and spatially can be responded to through eventual inclusion of Kantian forms of senses—time and space in the model. At the same time, given the nature of categories, they apply over any judgment around temporal/spatial relations as well—All A is on B, Some A precede B etc.

Achourioti & van Lambalgen (2011) in their work claim that their formalization of Kant's logic proves that "Kant's Table of Judgements…, is indeed, as Kant claimed, complete for the kind of semantics he had in mind". But the exhaustivity of the Kantian categories is not important for the work presented in this thesis. Rather, it is their formulation. As discussed in the previous section, the goal Kant had in the formulation of these categories was to give constitutive operations of the process of construction of the object of cognition. This along with the match between the properties of VSA and mind as cognized by Kant is what makes conception like Kantian categories important. I discuss this further in next section.

6.3.4 Kant and Vector Space Architecture

In this section, I discuss the points of concurrence between Vector Space Architecture and Kant's ontology of mind. A central thesis of Kant's work is a unified representation of knowledge of all objects (which he calls X) that we experience.

> "We find, however, that our thought of the relation of all cognition to its object carries something of necessity to it [. . .] since insofar as they are to relate to an object our cognitions must also necessarily agree with each other in relation to it, i.e., they must have that unity that constitutes the concept of an object. It is clear, however, that since

> we have to do only with the manifold of our representations, and that X which corresponds with them (the object) [. . .] is nothing for us, the unity that the object makes necessary can be nothing other than the formal unity of the consciousness in the synthesis of the manifold of representations"[8].

This formulation fits squarely with the global storage of information in a high-dimensional space which ensures that all representations occur in relation to one another. Kant wanted to explain how we construct the appearances so that our experience of the world is not a rhapsody of perception, but a unified experience with cognitively constructed objects. Instead of looking at a unified world to explain this, he turned inwards to unifying mechanism of the mind. The models of Vector Space architecture face the same issue now— how to encode highly complex, unorganized information of multiple modalities and render them meaningful in the intrinsically unified structure of its geometry. Kant's serious and extensive philosophical work in might have the right insights which are still being explored by cognitive science and has only recently been looked at for a theory of semantics. A complete application of Kant's work in this novel architecture is a project which will take a while to exhaust, especially given the notorious complexity of Kant's work. This thesis discusses the first steps towards this bigger project.

---

[8] Critique of Pure Reason, A104–105

6.4 Chapter Conclusion

What we find in Kant's philosophy is that it fully embraces the cognitive nature of our experience of the world. It thus departs from formal-logic semantics which treats its objects as non-psychological and attempts to relate them through set-theoretic relational forms. Kant's psychological leanings make his philosophy and the logical system presented therein are particularly suited for the goals of this thesis.

Chapter 7: Kantian HDM in R

7.0 Chapter Overview

The Chapter presents the system I created (called Kantian-HDM and created in R programming language) to implement Kantian categories in vector space architecture. One its novelties is the expansion of a feature of DSHM (*cardinal vectors*) to represent the Kantian categories. I first discuss how encoding and recall of information occurs in the system. I then proceed to perform basic information recall and similarity tests it on it. Post this basic testing, I discuss the positive inferential abilities that the system exhibits including the derivation of some of the non-critical categorial relations, analogical reasoning and the conjunction fallacy.

7.1 Cardinal Vectors

While vector space models have traditionally assumed no innate structure, by introducing *cardinal vectors* as universal atomic items, DSHM made a distinctly Kantian move. A key manner in which DSHM deviates from other VSA models is that the modelers added innate structure to memory. Rutledge-Taylor et al. (2014) found it "convenient to create universal atomic items, called cardinal items that represent special concepts" to "account for a variety of mental representations", such as *true* and *false,* or *good* and *bad,* in order to capture human behavior.

While for DSHM, the *cardinal vectors* were something created specifically as features for models of specific tasks, I created a system called Kantian-HDM (Arora & West, 2018) in the programming language R which makes *cardinal vectors* a core element of the model. The model is published on GitHub and is made open access. *Cardinal categories* in K-HDM encode the

Kantian categories and are consistently used to encode all conceptual relationships. This satisfied their universally necessary application which Kant had in mind when he conceptualized them.

Traditional semantic space models would treat the following three relationships—*Sun heats earth, Dogs make noise, Love conquers all* —as entirely distinct. Yet, understood through Kantian *categories*, these relationships share the same Kantian category types (Table 6). Any relationship can be described by the selection of one type from each of the four *categories* giving us a total of 3^4=81 basic relationships. For a few more examples, see Table 7.

Table 6: Kantian Categorization

Information	Relationship
Sun heats earth **Dogs make noise** ***Love conquers all***	Quantity**:** Universal
	Quality**:** Affirmative
	Relation**:** Hypothetical
	Modality**:** Assertoric

Table 7: Examples of Cardinal Categories for Different Relations

Relationship	Quantity	Quality	Relation	Modality
All bachelors are unmarried	Universal	Affirmative	Categorical	Apodeictic
All dogs are animals	Universal	Affirmative	Categorical	Assertoric
All flu is either viral or bacterial	Universal	Affirmative	Disjunctive	Apodeictic
Jack is stupid	Singular	Affirmative	Categorical	Assertoric
Mary makes cake	Singular	Affirmative	Hypothetical	Assertoric
Bill is not a doctor	Singular	Negative	Categorical	Assertoric
Some coma patients could be conscious	Particular	Affirmative	Categorical	Problematic
All Apples are red	Universal	Affirmative	Categorical	Assertoric

7.2 Storing information in K-HDM

7.2.1 Vectors in K-HDM

Like DSHM, K-HDM also represents concepts through two vectors—environmental and memory. An environmental vector ($\boldsymbol{E_c}$) which represents the percept of the concept, and which can be understood as its referents. A memory vector ($\boldsymbol{M_c}$) which encodes its relationship with other concepts. A concept has meaning as a result of these relationships, and thus $\boldsymbol{M_c}$ can be understood as the sense of the concept. Unlike concepts, the cardinal *categories* do not undergo a change in meaning as more knowledge is gathered. Thus, they will have one environment vector and no memory vector. This is because they are primitives and represent invariable pure concepts of understanding.

7.2.2 Encoding in K-HDM

All relations in K-HDM are stored as a subject–relation–object predication which is often algebraically written as xRy, where x is the subject, y is the object and R is the relation. Following this scheme would lead the memory chunks in DHSM to be consistently structured through same recursive operations. So, information such as "*some doors are red"* will have a memory chunk like that shown in.

Table **8**.

Table 8: Memory Chunk in Kantian HDM

Slot	Value	Environment Vectors	Memory Vectors
Subject	door	E_{door}	M_{door}
Predicate	red	E_{red}	M_{red}
Quantity	Particular	E_{uni}	–

Quality	Affirmative	E_{aff}	–
Relation	Categorical	E_{cat}	–
Modality	Assertoric	E_{asr}	–

All of the *categories* need to be bound first to form one of the unique 81 relationships; and so, we will have to create a *Relationship vector* which encodes the unique relationship through convolution of the respective cardinal vectors. Using this vector, information about subject and predicate is encoded in the memory vectors. For example, for information like "door is red" we create the relationship vector and update the memory vectors for *red* and *door* as given in Equation 4.

$$\mathbf{Rel}_{vec} = \mathbf{E}_{pat} * \mathbf{E}_{aff} * \mathbf{E}_{cat} * \mathbf{E}_{asr}$$

$$\mathbf{M}_{door} = \text{Old } \mathbf{M}_{door} + (\Phi * \mathbf{Rel}_{vec}) * \mathbf{E}_{red}$$

$$\mathbf{M}_{red} = \text{Old } \mathbf{M}_{red} + (\mathbf{E}_{door} * \mathbf{Rel}_{vec}) * \Phi$$

$$\mathbf{E}_{P1} = \mathbf{M}_{P1} = (\mathbf{E}_{door} * \mathbf{Rel}_{vec}) * \mathbf{E}_{red}$$

Equation 4: Storing information in K-HDM

Unlike DSHM, K-HDM stores all xRy relationships as new concepts with their own environmental and memory vectors. So "some doors are red" (let's call it concept P1) gets stored as a new concept. In this form, it is no longer a proposition, but a complex concept "some red doors". The initial value of the vectors for the concept is created by convolving the environmental vectors of all elements involved. This new concept can then be used to create more complex concepts through the same processes by which it was encoded.

Complicated relations such as “the sun heats the earth” can be encoded through these *categories* as well. The relationship is understood as containing two simpler relationships—"[the sun] *cause*s [[the earth] is [hot]]”. The categorical relation marks the proposition “the *earth is hot”* and is encoded using a Singular, Affirmative, Categorical and Assertoric combination of *categories.* The concept of “the hot earth” so created in then bound to the concept of “the sun” using a Singular, Affirmative, Hypothetical and Assertoric combination of *categories.* Together this encodes the essence of “the sun heats the earth” which is that the subject “the sun”, causes/results in a condition where “the earth is hot”.

7.3 Testing K-HDM

The model was tested in multiple ways and configurations over sparse datasets. The results from the same are discussed below. Keeping in mind the distinctions made in section 5.1 (pg. 56) regarding architectural, sub-symbolic and symbolic features, the tests were specifically intended to ensure understanding of the performance across all these levels.

7.3.1 Basic Encoding and Recall of Information (Architecture Level)

The dataset used for basic testing was a sparse dataset with 12 rows around features of Dogs and Cat (

Table 9). Of all the four predicates related affirmatively with *Dogs,* three were also related to *Cats* and two were related to *Birds*. This was confirmed during similarity recall where the model successfully demonstrated higher similarity between *Dogs* and *Cats* (62%) than *Dogs* and *Birds* (35%).

Table 9: Basic Dataset

Details	Subject	Quantity	Quality	Relation	Modality	Predicate
dogs are smart	Dogs	Universal	Affirmative	Categorical	Assertoric	Smart
dogs are sleepy			Affirmative			Sleepy
dogs are loyal			Affirmative			Loyal
dogs are big			Affirmative			Big
cats are smart	Cats		Affirmative			Smart
cats are sleepy			Affirmative			Sleepy
cats are not loyal			Negative			Loyal
cats are big			Affirmative			Big
birds are smart	Birds		Affirmative			Smart
birds are sleepy			Affirmative			Sleepy
birds are not loyal			Negative			Loyal
birds are not big			Negative			Big

In the test for demonstration of the fan-effect, I ran 20 simulations to encode only the rows where Quality category was Affirmative. The concept *Dogs* which has four such associations has a fan of four. Similarly, *Cats* and *Birds* have a fan of three and two respectively. For the fan effect to be demonstrated, the model should show a decrease in the average similarity between the concept and its closest vectors as the size of fan increases. Hence, the average similarity for *Dogs* should be the lowest and that for *Birds* be the highest. This is indeed what I found (Figure 16). As discussed earlier, given that K-HDM is a Vector Space model, it is no surprise that like DSHM and HDM, it also shows the fan-effect.

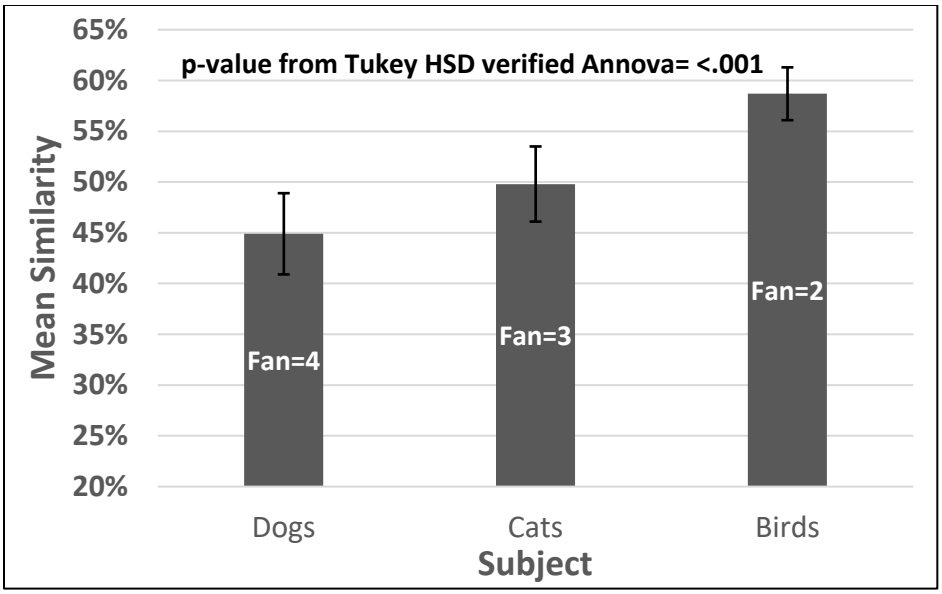

Figure 16: Fan Effect in K-HDM

7.4 Inference Behavior in K-HDM

7.4.1 Repeated storage and derivation of non-critical category types (Symbolic Level)

As discussed earlier in Criticality of Categories 6.3.2 (pg. 86), some of the Kantian categories seem to be more critical from the conceptual point of view than others. K-HDM has multiple control parameters one of which is “Repeated Storage”. This parameter allows for a control over whether repeated exposure to same proposition, e.g., ‘Dogs are loyal’ results in repeated storage of the relationship. This is a symbolic-level feature. Turning on the feature results in an interesting property where the model does not require explicit storage of the some of the Quantity” category types (*Universal* and *Particular*) but can rather derive it out of singular relationship stored as *Singular*. This is possible because the memory vector for a concept keeps a fuzzy count of the number of times that a relation has occurred. An estimate of this count can be derived by taking a projection of memory vector onto the vector corresponding to the relation.

A projection of a vector on another vector is a measure of the magnitude of the latter in the direction of former. This can be understood as the length of shadow that a vector would cast on another vector (Figure 17). Mathematically, projection of vector A on vector B can be computed by multiplying the magnitude of A with the cosine of the angle between A and B.

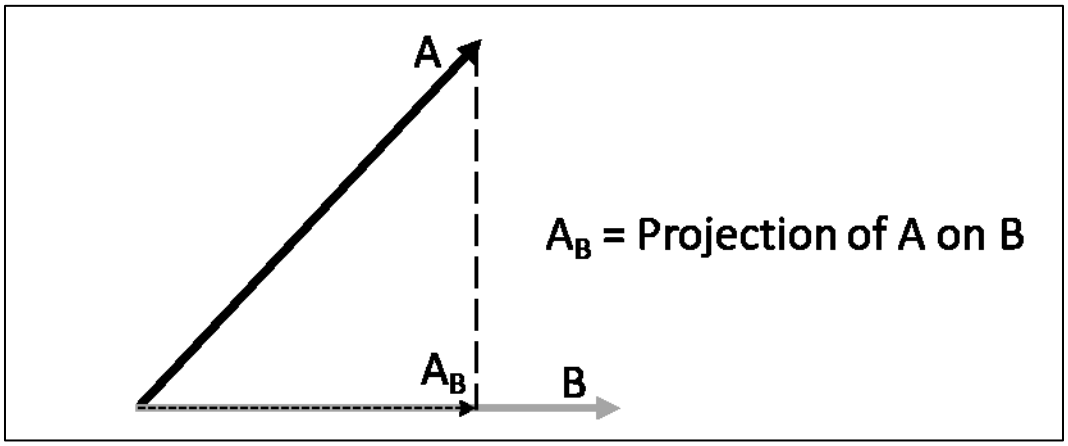

Figure 17: Dot Product as Projection of a vector onto another

For the simulation to demonstrate this, I generated 20x10 datasets containing randomly assigned status of 10 balls in a container. The balls are either broken or not broken. For every broken ball I encounter, I store "Ball is Broken" and for an unbroken ball, I store "Ball is not broken". At the end of the storage, a ratio of the projections (dot product) of memory vector of *Ball* on the vector representing "is broken" and "is not broken" allows me to estimate the actual ratio (see Equation 5). This is used to derive the quantity relationships. Figure 18 shows results averaged over 20 such simulations along with the error margin.

Equation 5: Estimate of Ratio in VSA

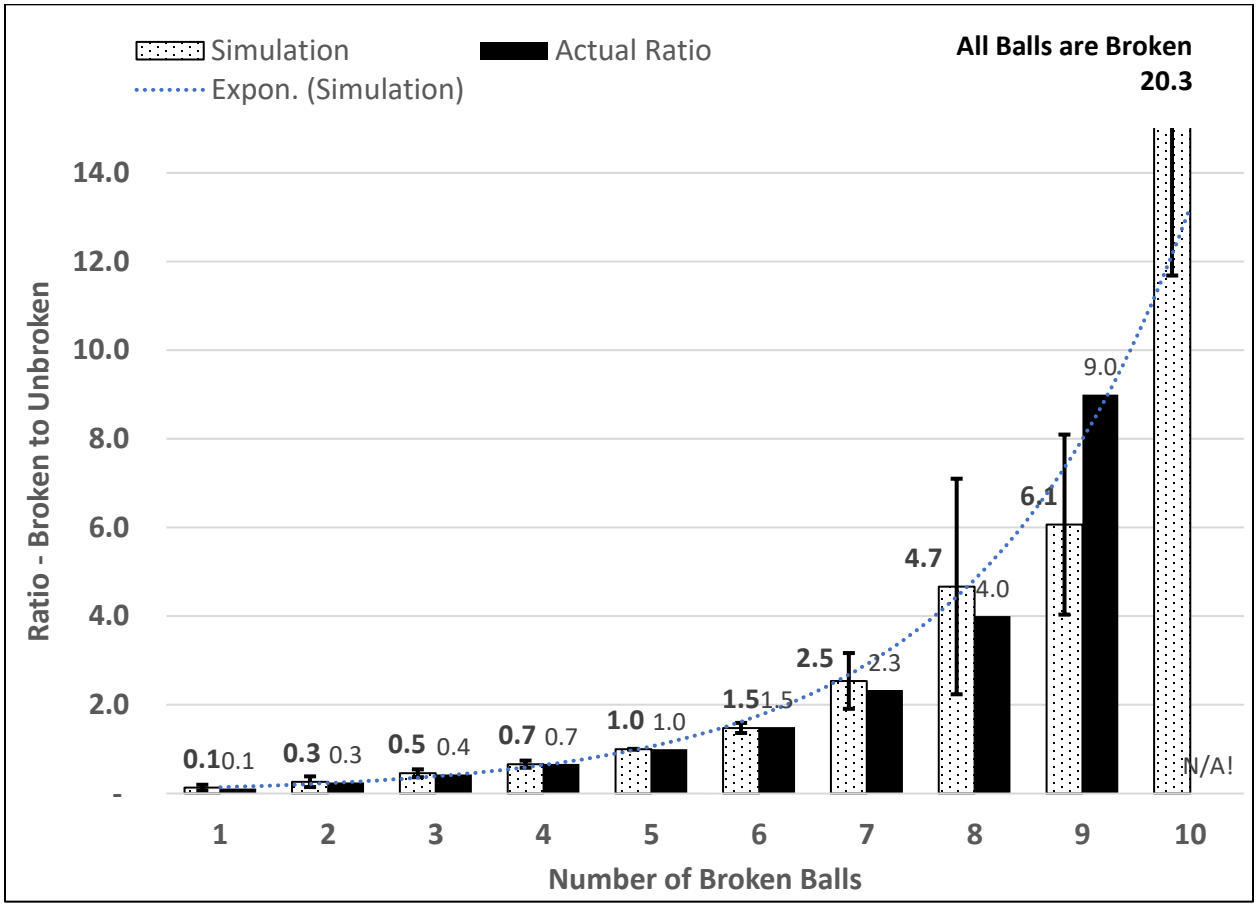

Figure 18: Projection Ratio and Quantity estimates

The estimated ratio grows in a pattern similar to the actual ratio of not broken to broken. The correlation between the actual ratio and projection ratio is quite high as well (R=0.96, p-value=<.001). The trendline in the graph demonstrates the exponential nature of the result. A point to note is the sudden jump at the last data point representing ten broken balls and zero unbroken balls. The actual ratio for this condition is infinite and non-computable because of divisibility with zero. The mean estimated ratio shows a huge jump for this condition (from 6.1 to 20.3). Such jumps are sufficient flags for the model to determine an extreme condition “All Balls

are Broken" for the model. The estimated ratio thus allows for an inference over the quantities even when no count is explicitly maintained in the system.

7.4.2 Analogical Reasoning

We discussed earlier that recalled vectors in VSA are noisy, but the robustness of the hyperspace allows for successful recall nevertheless. In addition to that, it is interesting that the noisy approximations of the recalled vector can also be used to solve proportional analogy problems of the form '*a* is to *b* as *c* is to ?'. Through this, the concepts are identified on the basis of shared structural relations. The noisy retrieval can be made more accurate by adding cues. For example, if we have stored "cold causes shiver" and "smoking causes cancer"; the analogical retrieval for *cold* is to *shiver* as *smoking* is to *?* would look as follows:

$$\text{Cue} = \mathbf{M}_{\mathbf{cold}} \oslash \mathbf{E}_{\mathbf{shiver}}$$

$$\mathbf{M}_{\mathbf{cancer}} \approx \text{recall vector} = \text{Cue} * \mathbf{E}_{\mathbf{smoking}}$$

The recall vector is usually found to be sufficiently similar to the memory vector of cancer ($\mathbf{M}_{\mathbf{cancer}}$) to successfully retrieve the latter. This similarity can be boosted by adding cues from other causal relationships stored in the system. For example, when the K-HDM system was trained on a database which contained many causal relationships like "smoking causes cancer", "bulb causes light", etc. and was given a query like "cold is to shiver, bulb is to light, as smoking is to ?". The query got boosted due to other causal relations stored in the system. For example, a boosted cue would be as follows:

$$\text{Cue} = \mathbf{M}_{\mathbf{cold}} \oslash \mathbf{E}_{\mathbf{shiver}} + \mathbf{M}_{\mathbf{bulb}} \oslash \mathbf{E}_{\mathbf{light}} + \ldots\ldots$$

The higher the number of cues, the more is the similarity between the recall vector resulting from convoluting the cue with the environmental vector for *smoking* and the memory vector for *cancer.* (Figure 19). We thus see the ability to derive proportional analogical inferences in K-HDM.

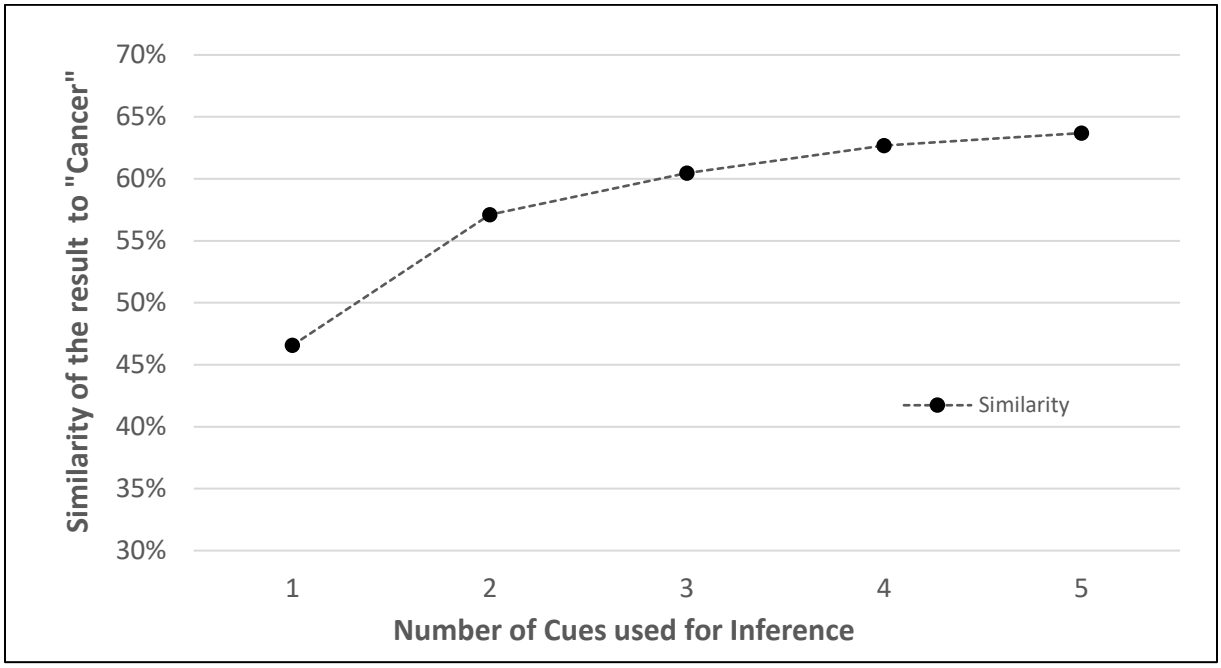

Figure 19: Cue Superposition Amplifies Analogical Retrieval

7.4.3 Quantum Probability

In this section, I discuss the possibility of explaining some of the strange ways in which human behave when making probabilistic decisions through the geometric operations of vector space. A detailed discussion of the same can be found in (Busemeyer, Pothos, Franco, & Trueblood, 2011).

In their famous experiment, Tversky & Kahneman, (1983) presented participants with a story about a hypothetical person, Linda, whose characteristics aligned her towards certain

character traits more than others. Participants were then asked to evaluate the probability of statements about Linda. The important comparison concerned the statements "Linda is a bank teller" (extremely unlikely given Linda's description) and "Linda is a bank teller and a feminist." Most participants chose the second statement as more probable than the first implying that for them:

Prob (bank teller) < **Prob** (bank teller & feminist)

This is not compatible or explicable through classical probability theory but falls out from quantum probabilistic theorization. Classical probability is a set-theoretic way to assign probabilities to the possible outcomes of a question. First, a sample space is defined, in which specific outcomes about a question are subsets of this sample space. Then, a probability measure is postulated, which assigns probabilities to disjoint outcomes in an additive manner (Kolmogorov, 1933/1950). The formulation is different in Quantum Probability theory, which is a geometric theory of assigning probabilities to outcomes. Under the theory, the probability of something is represented by the size of the projection of the state vector (a vector with all information about something) onto the vector representing a particular state. Such formulation allows for behavior like conjunction fallacy to emerge. For example (Figure 20), if we have a vector for *Linda*, and take a projection onto the vector for *Bank Teller*, we would expect to get a very small shadow ($\mathbf{OL_B}$) indicating a low probability of Linda being a bank teller. The projection of Linda onto vector representing *Feminist* though is much bigger ($\mathbf{OL_F}$). To get a vector for 'Linda is a feminist and a Bank teller', we take a projection $\mathbf{OL_F}$ onto **B** to get $\mathbf{OL_{FB}}$. As we can see in the diagram, $\mathbf{OL_{FB}}$ is greater than $\mathbf{OL_B}$ indicating a higher probability.

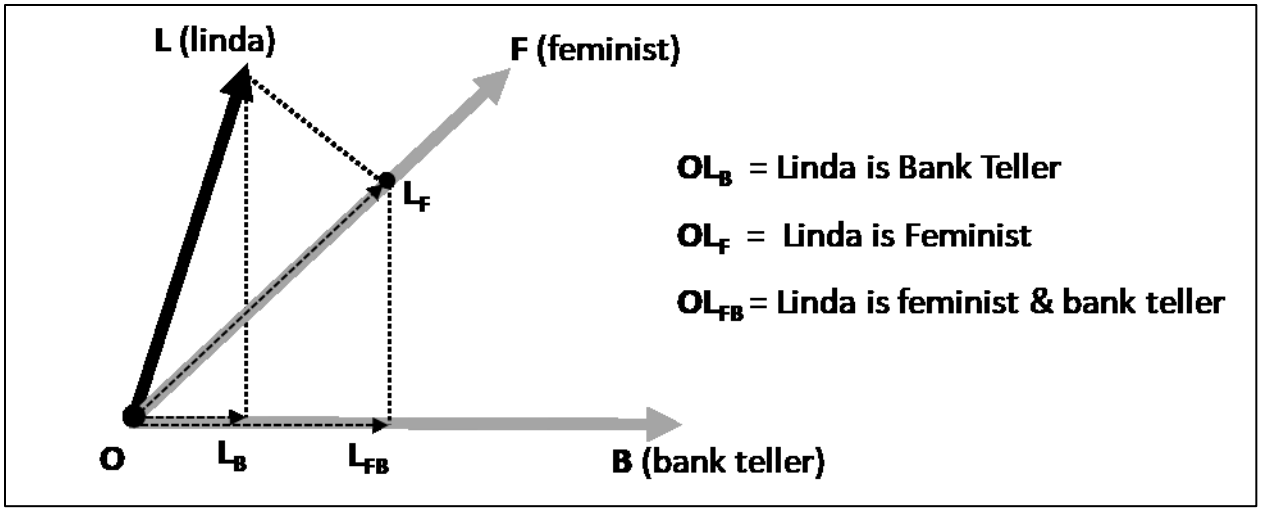

Figure 20: Computing Geometric Probabilities

To test such a behavior in K-HDM, I conducted 100 simulations which encoded initial information about concepts of Linda, Bank Teller, and Feminist. Each of three concepts were given four separate predicates unique to each of them. Along with that, each pair of concepts had some common predicates unique to each pair (Table 10).

Table 10: Unique and shared predicates for the concepts

	Linda	Feminist	Bank Teller
Linda	4	-	-
Feminist	10	4	-
Bank-Teller	3	6	4

The encoding was done in a way that it resulted in Linda being more similar to Feminist and hardly similar to Bank Teller, and the concept feminist is somewhat although not very similar to bank Teller (Figure 21). Then projections were computed to reflect the probability of Linda being a Bank Teller and Linda being both feminist and a Bank Teller.

The results show agreement with those presented in Tversky & Kahneman (1983). The projection of Linda for Bank Teller is less than that for Feminist and Bank Teller. We thus see that

K-HDM model does not just demonstrate general information encoding, but also exhibits some of the specific ways in which human infer information in practice which differs from how inferences occur in classical logic and probability.

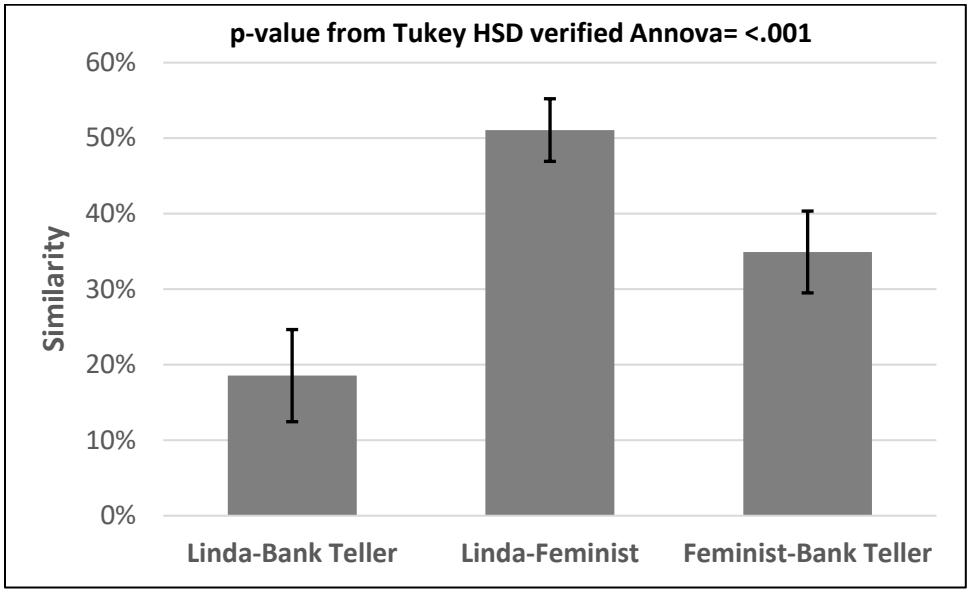

Figure 21: Similarity b/w *Linda*, Feminist, and Bank Teller

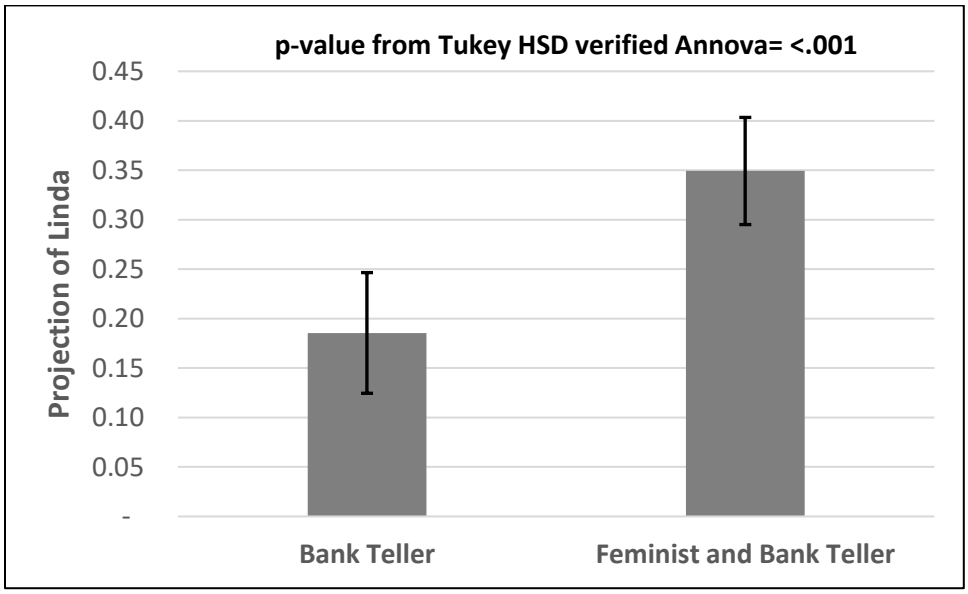

Figure 22: Projection of *Linda* Vector

7.5 Chapter Conclusion

The chapter presented a new model (K-HDM) specifically designed to encode information in xRy format where R is a unique relationship defined by Kantian categories implemented as cardinal vectors. x and y are the concepts which are being related. While successfully demonstrating general information encoding, the model also demonstrates inferential behavior which includes inferring quantity, analogical reasoning and cognitive effects over inference like the conjunction fallacy. Future works with regard to this model is discussed in the next section.

7.6 Conclusion and Future Works

The overall goal which motivated this work is to have a model of meaning which can satisfy the many constraints—computational, linguistics, psychological and neural—that any cognitive model of meaning must satisfy. The work of this thesis is only a part of this overarching project. The thesis identifies the need for organizational operations which systematically structure conceptual knowledge as an important part of this enterprise and proposes a model to do the same. Given the ambitious goal, the work in this thesis had to be conducted at multiple levels. With respect to analysis, it presents both methodology and implementation, and is both discovering and discursive. With respect to discipline, it is at the same time philosophical, computational and psychological.

The first requirement for any modeling process is to understand what it is that a model must accomplish in order to be able to successfully achieve its goal. This concern motivates chapter 1 and 2. These chapters are primarily philosophical in nature and attempt to delineate clearly the linguistic and psychological aspects of meaning and develop a body of constraints

through this analysis. Chapter 3 uses the products of these two chapters to analyze the existent models of meaning.

The second requirement for a modeling process is to identify the medium through which the modeling will be done. Hence, in Chapter 4, I make a case for vector space architecture to be able to provide the base system upon which models of meaning might be able to satisfy these constraints. Because VSA models suffer from a lack of clarity with regard to how they work and how the models should be assessed, chapter 5 proposes a hierarchical order through which VSA models can be designed and assessed. Secondly, I discuss the results from simulations I ran to shed light on what makes these highly mathematical, geometric systems works and discuss the reasons for the behavior which evolves from encoding information in them.

Thirdly, a modeling process requires a method through which it achieves its goal in the selected medium. For this, I turn towards Kantian philosophy in chapter 6 and make a case for why the psychological nature of the logic presented in it makes it the right choice of logical forms through which knowledge can be organized in VSA.

Finally, a modeling process must model and evaluate results. This is done in chapter 7 where I present the model I developed (Kantian-HDM) and the results I obtained with regards to the inference-making over the information encoded in it.

As with most interdisciplinary projects, a lot of work needs to go into developing the right interfaces such that the components of the system hold together. Complete integration of Kant's categories into vector space models is an ongoing project of which this thesis makes the first

concrete steps. As I continue investigating the implementation of categorial structure in VSA models, the next step that I am working towards is integrating a perceptual VSA model with this system. Further work is also needed in establishing the right connections between concepts like action, processes, etc., with Kantian categories. Kant mentions that such explication of categories and combination with spatial and temporal forms will give us derivative *a priori* concepts like force, action, undergoing, presence, etc.[9] But apart from a mere mention, he does not give any details.

There is much left to discover in the rich mine that Kant's collected works is and the same can be said about geometric intricacies of vector space architecture. I hope this thesis has at least made outlined the potential that the merger of the two shows towards developing a cognitive theory of human semantics.

[9] Critique of Pure Reason, A82, B108

References

Achourioti, T., & van Lambalgen, M. (2011). A formalization of Kant's transcendental logic. *The Review of Symbolic Logic*, *4*(2), 254–289. https://doi.org/10.1017/S1755020310000341

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology*, *6*, 451–474. https://doi.org/10.1016/0010-0285(74)90021-8

Anderson, J. R., & Lebiere, C. (1998). *The Atomic Components of Thought*. Mahwah, NJ: Lawrence Erlbaum Associates.

Antony, L., & Davies, M. (1997). Meaning and semantic knowledge. *… of the Aristotelian Society, Supplementary Volumes*, *71*(1997), 177–209. Retrieved from http://www.jstor.org/stable/10.2307/4106958

Arora, N., & West, R. (2018). *Kantian Holographic Declarative Memory Module in R*. https://doi.org/10.5281/ZENODO.1217790

Bannert, M. M., & Bartels, A. (2013). Decoding the yellow of a gray banana. *Current Biology*, *23*(22), 2268–2272. https://doi.org/10.1016/j.cub.2013.09.016

Baroni, M. (2013). Composition in distributional semantics. *Linguistics and Language Compass*, *7*(10), 511–522. https://doi.org/10.1111/lnc3.12050

Barton, R. A. (2012). Embodied cognitive evolution and the cerebellum. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1599), 2097 LP-2107. Retrieved from http://rstb.royalsocietypublishing.org/content/367/1599/2097.abstract

Bauer, A. J., & Just, M. A. (2015). Monitoring the growth of the neural representations of new animal concepts. *Human Brain Mapping*, *36*(8), 3213–3226. https://doi.org/10.1002/hbm.22842

Behaghel, O. (1932). *Deutsche Syntax: eine geschichtliche Darstellung. Wortstellung, Periodenbau*. Winter. Retrieved from https://books.google.ca/books?id=1kdTAAAAcAAJ

Berwick, R. C. (1989). Learning word meanings from examples. *Semantic Structures. Advances in Natural Language Processing*, 89–124.

Brook, A. (1994). *Kant and the Mind*. Cambridge University Press. Retrieved from https://books.google.ca/books?id=e-Jhd6cUe5IC

Burgess, C., & Lund, K. (1997). Modelling parsing constraints with high-dimensional context space. *Language and Cognitive Processes*, *12*(2), 177–210. https://doi.org/10.1080/016909697386844

Burgess, C., & Lund, K. (2000). The dynamics of meaning in memory. *In Cognitive Dynamics: Conceptual Change in Humans and Machines. Dietrich & Markman (Eds.), Psychology Press*, (May), 117–156. https://doi.org/10.1016/j.cognition.2007.07.015

Busemeyer, J. R., Pothos, E. M., Franco, R., & Trueblood, J. (2011). A quantum theoretical explanation for probability judgement errors. *Psychological Review*, *118*, 193–218. https://doi.org/10.1037/a0022542

Chalmers, D. J., French, R. M., & Hofstadter, D. R. (1992). High-level perception, representation, and analogy: A critique of artificial intelligence methodology. *Journal of Experimental and Theoretical Artificial Intelligence*, *4*(3), 185–211. https://doi.org/10.1080/09528139208953747

Chertkow, H., Bub, D., Deaudon, C., & Whitehead, V. (1997). On the Status of Object Concepts in Aphasia. *Brain and Language*, *232*(58), 203–232.

Chomsky, N. (1957). *Syntactic Structures*. Bod Third Party Titles. Retrieved from https://books.google.ca/books?id=a6a_b-CXYAkC

Chomsky, N. (1980). Rules and representations. *Behavioral and Brain Sciences*, *3*(127).

Chomsky, N. (1988). *Language and Problems of Knowledge: The Managua Lectures*. Cambridge University Press. Retrieved from https://books.google.ca/books?id=hwgHVRZtK8kC

Chomsky, N. (2014). Minimal recursion: Exploring the prospects. In *Recursion: Complexity in cognition* (pp. 1–15). Springer.

Clarke, A., & Tyler, L. K. (2015). Understanding What We See: How We Derive Meaning From Vision. *Trends in Cognitive Sciences*, *19*(11), 677–687. https://doi.org/10.1016/j.tics.2015.08.008

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*(2), 240–247. https://doi.org/10.1016/S0022-5371(69)80069-1

Damasio, H., Tranel, D., Grabowski, T., Adolphs, R., & Damasio, A. (2004). Neural systems behind word and concept retrieval. *Cognition*, *92*(1–2), 179–229.

de Vega, M., Glenberg, A. M., & Graesser, A. C. (2008). *Symbols and Embodiment: Debates on Meaning and Cognition*. Oxford University Press. Retrieved from https://books.google.ca/books?id=D_PWAAAAMAAJ

Dever, J. (1999). Compositionality as methodology. *Linguistics and Philosophy*, *22*(3), 311–326.

Donnellan, K. S. (1972). Proper names and identifying descriptions. In *Semantics of natural language* (Vol. 21, pp. 356–379). Springer.

Eagleman, D. M. (2008). Human time perception and its illusions. *Current Opinion in Neurobiology*, *18*(2), 131–136. https://doi.org/10.1016/j.conb.2008.06.002

Franklin, D. R. J., & Mewhort, D. J. K. (2015). Memory as a hologram: An analysis of learning and recall. *Canadian Journal of Experimental Psychology*, *69*, 115–135. https://doi.org/10.1037/cep0000035

Gainotti, G. (2011). The organization and dissolution of semantic-conceptual knowledge: Is the "amodal hub" the only plausible model? *Brain and Cognition*, *75*(3), 299–309. https://doi.org/10.1016/j.bandc.2010.12.001

Gayler, R. W. (2003). Vector symbolic architectures answer Jackendoff's challenges for cognitive neuroscience. In P. Slezak (Ed.), *Proceedings of the Joint International Conference on Cognitive Science* (pp. 133–138). Sydney, Australia: University of New South Wales.

Glaser, W. R. (1992). Picture naming. *Cognition*, *42*(1–3), 61–105. https://doi.org/10.1016/0010-0277(92)90040-O

Glenberg, A. M., & Mehta, S. (2012). The limits of covariation. In *Symbols and Embodiment: Debates on Meaning and Cognition*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199217274.003.0002

Hagoort, P., Baggio, G., & Willems, R. M. (2009). Semantic Unification. In *The cognitive neurosciences, 4th ed.* (pp. 819–836). MIT press. https://doi.org/10.1002/int.4550070108

Hanna, R. (2017). Kant's Theory of Judgment. In *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/win2017/entries/kant-judgment/

Hauser, M. D., Chomsky, N., & Fitch, W. T. S. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science*, *298*(5598), 1569–1579. https://doi.org/10.1126/science.298.5598.1569

Isac, D., & Reiss, C. (2008). General Requirements on Grammars. In *I-Language: An Introduction to Linguistics as Cognitive Science*. OUP Oxford.

Jackendoff, R. (1983). *Semantics and Cognition*. de Gruyter. Retrieved from https://books.google.ca/books?id=a0fQYPHBc4gC

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press.

Jackendoff, R. (2007a). A Parallel Architecture perspective on language processing. *Brain Research*, *1146*(1), 2–22. https://doi.org/10.1016/j.brainres.2006.08.111

Jackendoff, R. (2007b). *Language, Consciousness, Culture: Essays on Mental Structure*. *The Jean Nicod Lectures* (Vol. 2007). Bradford Book. https://doi.org/10.1017/CBO9781107415324.004

Jackendoff, R. (2017). In Defense of Theory. *Cognitive Science*, *41*, 185–212. https://doi.org/10.1111/cogs.12324

Jackendoff, R., & Pinker, S. (2005). The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, and Chomsky). *Cognition*, *97*(2), 211–225. https://doi.org/10.1016/j.cognition.2005.04.006

Johnson-Laird, P. N. (1982). Formal Semantics and the Psychology of Meaning. In S. Peters & E. Saarinen (Eds.), *Processes, Beliefs, and Questions: Essays on Formal Semantics of Natural Language and Natural Language Processing* (pp. 1–68). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-015-7668-0_1

Johnson-Laird, P. N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Harvard University Press. Retrieved from https://books.google.ca/books?id=FS3zSKAfLGMC

Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, *114*(1), 1–37. https://doi.org/10.1037/0033-295X.114.1.1

Kachergis, G., Cox, G. E., & Jones, M. N. (2011). OrBEAGLE: Integrating orthography into a holographic model of the lexicon. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *6791 LNCS*(PART 1), 307–314. https://doi.org/10.1007/978-3-642-21735-7_38

Kant, I. (1781). *Critique of Pure Reason*. (W. S. Pluhar & P. Kitcher, Trans.). Hackett Publishing Company. Retrieved from https://books.google.ca/books?id=Iz1xiAIcWiMC

Kant, I., (Ed. Förster, E., & Rosen, M.) (1995). *Opus Postumum*. Cambridge University Press. Retrieved from https://books.google.ca/books?id=pJWKGVPjRbIC

Katz, J. J., & Fodor, J. a. (1963). The Structure of a Semantic Theory. *Language*, *39*(2), 170–210. https://doi.org/10.2307/411200

Kelly, M. A. (2016). The memory tesseract: Developing a unified framework for modelling memory and cognition. In *Curve Theses and Dissertations Collection*. Carleton University, Canada. Retrieved from https://curve.carleton.ca/9d053d1d-0bfb-4109-a9a1-70a34873014a

Kelly, M. A., & Reitter, D. (2017). Holographic Declarative Memory: Using Distributional Semantics within ACT-R. *AAAI Fall Symposium Series; 2017 AAAI Fall Symposium Series*. Retrieved from https://aaai.org/ocs/index.php/FSS/FSS17/paper/view/16001

Kelly, M. A., & West, R. L. (2012). From Vectors to Symbols to Cognition: The Symbolic and Sub-Symbolic Aspects of Vector-Symbolic Cognitive Models. *Proceedings of the 34th Annual Meeting of the Cognitive Science Society (CogSci 2012)*, 1768–1773. Retrieved from http://palm.mindmodeling.org/cogsci2012/papers/0311/paper0311.pdf

Kolmogorov, A. N. (2018). *Foundations of the Theory of Probability: Second English Edition*. Courier Dover Publications.

Kornai, A., & Kracht, M. (2015). Lexical Semantics and Model Theory: Together at Last? In *Proceedings of the 14th Meeting on the Mathematics of Language (MoL 2015)* (pp. 51–61). Stroudsburg, PA, USA: Association for Computational Linguistics. https://doi.org/10.3115/v1/W15-2305

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211–240.

Levelt, W. J. M. (1993). *Speaking: From Intention to Articulation*. Bradford Books,U.S. Retrieved from https://books.google.ca/books?id=LbVCdCE-NQAC

Levelt, W. J. M. (2013). *A History of Psycholinguistics: The Pre-Chomskyan Era*. *A History of Psycholinguistics: The Pre-Chomskyan Era*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199653669.001.0001

Mazzone, M. (2014). A Generative System for Intentional Action? *Topoi*, *33*(1), 77–85. https://doi.org/10.1007/s11245-013-9186-7

Mitchell, J., & Lapata, M. (2008). Vector-based Models of Semantic Composition. *Acl*, *8*(June), 236–244. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.164.9603&rep=rep1&type=pdf%5Cnhttp://homepages.inf.ed.ac.uk/s0453356/composition.pdf

Mitchell, J., & Lapata, M. (2010). Composition in distributional models of semantics. *Cognitive Science*, *34*, 1388–1429. https://doi.org/10.1111/j.1551-6709.2010.01106.x

Murphy, G. (2004). *The Big Book of Concepts*. MIT Press. Retrieved from https://books.google.ca/books?id=t2jldRsNkgsC

Nilsson, N. J. (1984). *Shakey The Robot*. 333 Ravenswood Ave., Menlo Park, CA 94025.

Padó, S., & Lapata, M. (2003). Constructing semantic space models from parsed corpora. *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - ACL '03*, *1*, 128–135. https://doi.org/10.3115/1075096.1075113

Partee, B. H. (2014). A Brief History of the Syntax-Semantics Interface in Western Formal Linguistics. *Semantics-Syntax Interface*, *1*(1), 1–21. Retrieved from http://semantics-syntax.ut.ac.ir

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*(12), 976. https://doi.org/10.1038/nrn2277

Pinker, S. (1994). *The Language Instinct (1994/2007)*. New York, NY: Harper Perennial Modern Classics.

Plate, T. A. (1995). Holographic reduced representations. *IEEE Transactions on Neural Networks*, *6*, 623–641. https://doi.org/10.1109/72.377968

Roeper, T., & Speas, M. (2014). *Recursion: Complexity in Cognition*. *Recursion: Structural Complexity in Language and Cognition* (Vol. 43). Springer International Publishing. https://doi.org/10.1007/978-3-319-05086-7

Roy, D. (2008). A mechanistic model of three facets of meaning. In *Symbols and Embodiment*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199217274.003.0011

Russell, B. (1948). *Human knowledge, its scope and limits*. Simon and Schuster. Retrieved from https://books.google.ca/books?id=4HoIAQAAIAAJ

Rutledge-Taylor, M. F., Kelly, M. A., West, R. L., & Pyke, A. A. (2014). Dynamically structured holographic memory. *Biologically Inspired Cognitive Architectures*, *9*, 9–32. https://doi.org/10.1016/j.bica.2014.06.001

Sellars, W. S. (1962). Philosophy and the scientific image of man. In R. Colodny (Ed.), *Science, Perception, and Reality*. Humanities Press/Ridgeview.

Sloboda, M. J. (1995). *Schemata as monogrammata: Opening the way towards a Kantian phenomenology of meaning*. *ProQuest Dissertations and Theses*. University of Toronto (Canada), Ann Arbor. Retrieved from http://proxy.library.carleton.ca/login?url=https://search.proquest.com/docview/304268925?accountid=9894

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Harvard University Press. Retrieved from https://books.google.ca/books?id=IfkMAQAAMAAJ

Smith, N. K. (1918). A commentary to Kant's "Critique of pure reason." London.

Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence*, *46*(1–2), 159–216. https://doi.org/10.1016/0004-3702(90)90007-M

Snyder, A., Bossomaier, T., & Mitchell, D. J. (2004). Concept formation: "object" attributes dynamically inhibited from conscious awareness. *Journal of Integrative Neuroscience*, *3*(1), 31–46. https://doi.org/10.1142/S0219635204000361

Strawson, P. F. (1959). *Individuals*. Retrieved from http://books.google.co.uk/books/about/Individuals.html?id=3uSO7aToYsAC&pgis=1

Treisman, A. M. (1998). Feature Binding, Attention and Object Perception. *Philosophical Transactions: Biological Sciences*, *353*(1373), 1295–1306. https://doi.org/10.1098/rstb.1998.0284

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*(4), 293.

Vinson, D. P. (2009). Representing meaning: a feature-based model of object and action words, 1–157. Retrieved from http://eprints.ucl.ac.uk/14891/

Widdows, D., & Cohen, T. (2014). Reasoning with vectors: A continuous model for fast robust inference. *Logic Journal of the IGPL*, *23*(2), 141–173. https://doi.org/10.1093/jigpal/jzu028

Williams, T. C. (1993). *Kant's Philosophy of Language Chomskyan Linguistics and its Kantian Roots*.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, *104*(19), 7780–7785.