Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 145 (2018) 680–690

Procedia Computer Science

www.elsevier.com/locate/procedia

Postproceedings of the 9th Annual International Conference on Biologically Inspired Cognitive Architectures, BICA 2018 (Ninth Annual Meeting of the BICA Society)

Why the Common Model of the mind needs holographic a-priori categories

Nipun Arora[a]*, Robert West[a], Andrew Brook[a], and Matthew Kelly[b]

[a]*Carleton University, 1125 Colonel By Dr, Ottawa, ON K1S 5B6, Canada*
[b]*The Pennsylvania State University, University Park, 16802 - 6823, USA*

Abstract

The enterprise of developing a common model of the mind aims to create a foundational architecture for rational behavior in humans. Philosopher Immanuel Kant attempted something similar in 1781. The principles laid out by Kant for pursuing this goal can shed important light on the common model project. Unfortunately, Kant's program has become hopelessly mired in philosophical hair-splitting. In this paper, we first use Kant's approach to isolate the founding conditions of rationality in humans. His philosophy lends support to Newell's knowledge level hypothesis, and together with it directs the common model enterprise to take knowledge, and not just memory, seriously as a component of the common model of the mind. We then map Kant's cognitive mechanics to the operations which are used in the current models of cognitive architecture. Finally, we argue that this mapping can pave the way to develop the ontology of the knowledge level for general intelligence. We further show how they can be actualized in a memory system using high dimensional vectors to achieve specific cognitive abilities.

Keywords: Common Model of Cognition; Kant; Reason; Holographic Memory; Knowledge Level

* Corresponding author. Tel.: +1-613-981-3942
E-mail address: nipun.arora@carleton.ca

10.1016/j.procs.2018.11.060

1. Introduction

Laird, Lebiere, & Rosenbloom (2017) [1] argue that research into understanding the human mind should be guided by a standard/common model of the mind (CMM). The common model is to serve as a foundation upon which higher-order cognitive abilities are realized through acquired knowledge and skills. Laird et al. propose a common model based on widespread consensus in the field regarding the fixed components of the mind (see Figure 1).

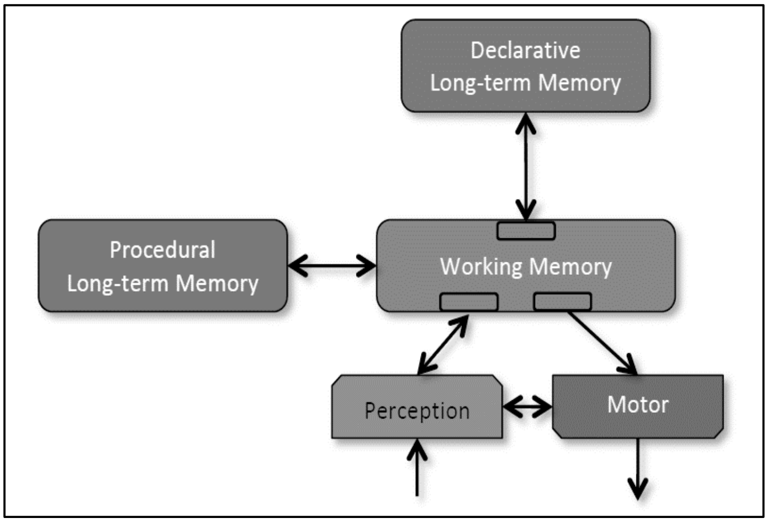

Figure 1: The standard model of mind (Laird et al., 2017)

There are three features of a common model of the mind that need to be emphasized—it must be integrative in nature, foundational in its cognitive role, and sufficient to give rise to rational behavior. It is integrative because the exercise of building a common model is not as much about making new empirical discoveries about the faculties of the mind, as it is about putting the existent theories of these faculties together. This is because the benefits of having a common model do not result from novelty in the components of the model itself, but rather arise because integration constrains the theories to lead to a parsimonious whole which is greater than the sum of its parts. The integration is novel; the components do not need to be. And hence, the hard questions here are the ones concerning the relation of the components with each other. Furthermore, the intent is for the common model to be foundational, and hence, it must be sufficiently general that all high-order cognitive activities are expressible in the primitive functions of this model. In that sense, such a model constitutes human intelligent behavior in its most foundational cognitive form. Finally, the question arises regarding the evaluation of the model's sufficiency. What basic cognitive achievement must the model have in order to be deemed minimally appropriate. Laird et al.[1] identify *bounded rationality* as a central feature of human-like minds which these models much achieve. Reason thus becomes the touchstone criterion that a common model of the mind must achieve. The question which then arises is this—*what are the conditions that constitute the ground for bounded rational behavior?*

While Laird et al. [1] cite many modern cognitive architectures, going further back in history, one inevitably encounters the infamously cryptic work of philosophy known as *The Critique of Pure Reason* [2] written by Immanuel Kant (1781) two centuries before the cognitive revolution began. Kant is regarded as the most important figure in the history of modern western philosophy and *The Critique of Pure Reason* is his magnum opus. To say that Kant's work is difficult to comprehend is an understatement. Philosophy of mind is still attempting to recover all of his insights about human reason which is a central theme of his work. After identifying Reason as a core human trait, and before giving his critique of it, Kant painstakingly discusses the question of *what the conditions that constitute the ground for rational behavior are*. This led him to theorize that the mind has the faculties of *Sensibility* (Perception), *Understanding,* and *Reason* (see Figure 2). These faculties form a highly interactive system that comes together to create a rational being. Centuries before Newell made a call for unified theories of cognition, Kant attempted to give such a theory. It is for this reason that Kant's work is relevant to the common model enterprise and his philosophy can serve as a guide in its further development.

Kant identifies objectivity of our experience as the condition for rationality and further claims that reason binds

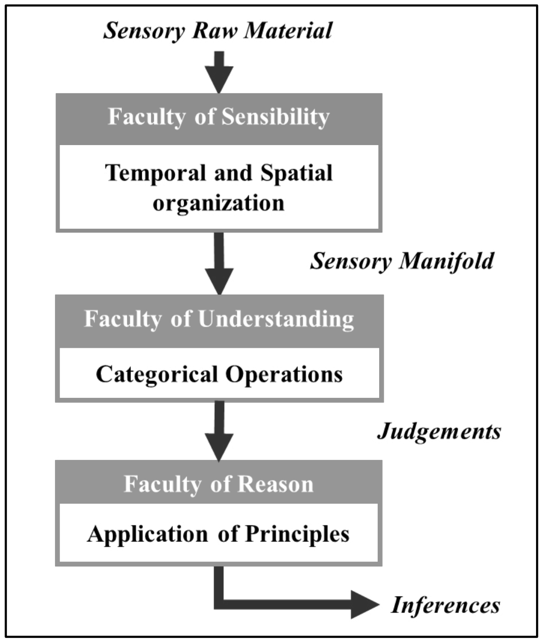

Figure 2: Kant's Unified Model

these experiences into a unified network of world knowledge. By objective Kant does not mean 'factual', but rather refers to the objectification of our subjective experience of the world such that we think in terms of objects and not merely the sensations we experience. What he is talking about here is the ability of our skull-bound brain to create from the subjective, isolated experiences of the world that we have, for example, 'the feeling of heaviness when picking a body', a reliable law-like objective judgment, for example, 'bodies are heavy'. Kant proposes integrative operations that he calls *a-priori categories*. These he claims are universal and necessary operations for producing an objective experience. He further proposes the principles of Reason which unify experiences to create knowledge.

Knowledge in current models of cognitive architectures is realized through the memory modules with ontologies that are specific to the tasks to be modeled and are unrelated to the rest of the cognitive system. Given the interdependence that Kant's philosophy imposes between Reason and knowledge-creation, it is important for the common model of the mind to consider the nature of knowledge necessary for human-like rational behavior. A Kantian approach thus affirms Newell's call for a distinct knowledge level in cognition [3], and further guides us in how this knowledge level fits into the design of a rational agent. In terms of the common model, the Kantian argument is that the mind requires some a-priori categories to produce objective-knowledge. Our claims, which we'll call the common model arguments, are different and go as follows.

- **Observation 1:** A theory of the conditions for rational behavior is important for the common model enterprise. Kant holds that rational behavior requires objective-knowledge and it is certain innate operations (a-priori categories) that allow us to create such knowledge.
- **Claim 1:** The typical operations in contemporary common model architectures like SOAR, ACT-R, etc., map onto these Kantian a-priori categories.
- **Observation 2:** Yet, unlike Kantian categories, these operations do not play a role in creating objective knowledge in these architectures. The operations merely act over pre-given data structures and are otherwise unrelated to knowledge represented in them.
- **Claim 2:** If Kant is right, in order to have general intelligence, we must have a knowledge level wherein objective-knowledge is realized via these operations. His a-priori categories, which are reflected in the operations

of cognitive architecture on one hand, also correspond to general concepts that structure our knowledge on the other hand. To have general human-like intelligence, models of rational agents require memory modules that have these a-priori categories as ontologies. We illustrate this claim with results from simulations run in a memory system based on a vector architecture that instantiates these categories.

2. Reason, Objective Knowledge, and a-priori Categories

Russell [4] describes Plato's problem as the following— "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?" The question digs deep into the issues of our knowledge of the world, and our ability to reason through and expand it. The relation between rational behavior and knowledge is one of the most contentious topics in AI. According to Newell [3], "Knowledge is whatever can be ascribed to an agent, such that the agent's behavior can be computed according to the principles of rationality.". Newell thus uses rational behavior as an evaluative criterion in order to give a functional definition to knowledge instead of a more traditional truth-based definition. This has been criticized by other AI philosophers who claim that "according to the modeling view, knowledge is much more related to the classical notion of truth intended as correspondence to the real world, and less dependent on the particular way an intelligent agent pursues its goals." [5,6]). This pushes the focus towards using the ontological truth about the world in order to develop ontologies for knowledge bases.

Kant's philosophy, almost in premonition, addresses this through an alternative, cognitive definition of Truth and Reason. For Kant, "Truth consists in the agreement of a cognition with its object"[†] where this object is cognitively constructed. Additionally, true cognitions must find a definite place within the single, unified knowledge of the world. Reason creates this unification and thus becomes "the ultimate arbiter of truth" so defined [7]. Through his account, Kant explicates that Reason is such an important cognitive feature because it optimizes an agent's cognition by judging its fitness with respect to a unified knowledge of the world, and truthiness of the cognition is this fitness.

A key claim here is that reason operates over objective-experience. The point, characterized by Clark [8] as the "view from inside the black box", is that the brain is essentially locked in a skull and separated from the world, yet we go beyond the mere subjective data of how we experience the world to having a law-like, objective knowledge of the objects in the world. We assume that there exists an object, and this hypothesis that there exists an object to which aspects of experience are referred is part of the perceptive process [9]. According to Kant, humans synthesize the objects of the world from a subjective experience. Hence, our knowledge corresponds to objects that are not platonic in their existence, but rather, reside in our minds, constructed cognitively, and designed to represent the world outside us. Kant claimed that this objective nature of knowledge is a result of the innate faculties of the mind. In other words, our knowledge and experience have the nature that they do not merely because of the world that they seem to refer to, but because of the mental faculties which we create them.

Kant called the operations of these faculties a-priori categories. The a-priori categories are universally and necessarily true of all objects because they reflect the innate processes that create these cognitive objects from the sensory information we receive. Table 1 gives the categories that, according to Kant, exhaustively capture the kinds of operations that structure our knowledge. To derive these innate categories, Kant first asks what kind of logical judgments we can have that describes relations between pairs of concepts. Kant derives a list of four categories, each with three types, making a total of twelve. Each of these twelve category-types manifests as a pure concept. For example, Hypothetical as Causality, Singular as Unity, Apodeictic as Necessity, etc. For Kant, the logical forms represent the central operations of the mind, while the pure concepts are their marks on our experience of the world.

[†] Critique of Pure Reason, A 59/B 83

Table 1. Table of Categories

Categories	Logical Judgement	Type	Pure Concept
	All A is B	Universal	All-ness
Quantity	Some A is B	Particular	Plurality
	A is B	Singular	Unity
	A is B	Affirmative	Reality
Quality	A is not B	Negative	Negation
	A is non-B	Infinite	Limitation
	A is B	Categorical	Inherence
Relation	If A, then B	Hypothetical	Causality
	A is B, C or D	Disjunctive	Community
	A may be B	Problematic	Possibility
Modality	A is B	Assertoric	Existence
	A must be B	Apodeictic	Necessity

3. **Common Model and a-priori Categories**

Most cognitive architectures, such as SOAR and ACT-R, have something akin to procedural and working memories (whether as separate modules or as part of a singular memory system). These control the system by containing and implementing condition-action pairs (if A, then B) where satisfaction of a criterion A leads to a subsequent result B. This requires a matching operation over the conditionals. The strictness of the scope of this match can vary such that it requires satisfaction of merely one, some, or all criteria. Finally, the system must also be able to resolve conflicts when there are multiple matching conditionals. In case of a single match, the rule is executed as is, but in the case of multiple matches, a selection must be made of which to execute. This 'single rule implementation' results in a serial bottleneck in performance. These features are reflected exactly in the Kantian categories.

The most important Kantian category is that of *Relation* which has the *categorical*, *hypothetical* and *disjunctive* types. Matching what is given in a working memory buffer to production rules is reflected in the *categorical* relation (Given = A). The production rules themselves are reflected in the *hypothetical* relation (If Given = A, then B). When more than one production rule matches, it is a *disjunctive* relation (If Given=A, then B or C or D). According to Kant, the disjunctive relation describes a choice, but also includes the idea that the options are mutually exclusive, i.e., only one can be actualized at a time. This results in a serial bottleneck that Laird et al. [1] recognize as a core feature of the cognitive act cycle.

For the *Quality* category, a match between a production and the buffers is an *affirmative* relation. But if the production matching condition is negated the match is *negative* or *infinite*, depending on the number of alternatives. The strictness of the scope of these matches is reflected in the *Quantity* category with its *universal*, *particular*, and *singular* types, which map onto the requirement that all, some, or a particular instance of a condition be satisfied for a match.

What makes the production rules themselves possible? This is where the *Modality* category comes into play. When one looks at the structure of production rules (If A, then B), we see that whereas A is considered merely as a possibility (thus *problematic*, see Table 1), B is considered an assertion (*assertoric*). In the case of multiple production rule matches, we have (as mentioned above) a *disjunctive* choice of assertions. In this case, further selection is made to select one. Moving from match to execution of a selected rule is reflected in the "*must*" of the *apodeitic type*. In fact, Kant alludes to this serial movement from *If* (production condition), to *Is* (matching), and finally to *Must* (execution) as "so many moments of thought as such"‡. In cognitive modeling terms, what he means is that different modalities reflect the stages of the cognitive cycle.

This close mapping between Kantian categories and operations in the common model hardly seems like a coincidence. The very operations that cognitive modelers postulate are needed in the common model are those that Kant found in his analysis of Reason. Furthermore, the claim that the common model should be a sufficient foundation for high-order cognition parallels Kant's claim that his categories are sufficient to create objective-knowledge (for a formal defense of the sufficiency of Kant's claim see Achourioti & van Lambalgen [10]).

4. Declarative Memory and a-priori Categories

Current models of cognitive agents lack general intelligence. We believe that the issue here is that these agents do not create their knowledge in a generalized manner. While most models of rational agents have similar general working-memory operations, the contents of the declarative memory in these agents and corresponding learning ontologies are structured for specific tasks. This is justified via a claim that high-order cognitive tasks are differentiated from each other by acquired skills and knowledge and not by the underlying cognitive architecture. But there must be something common across these specialized domains of knowledge due to their creation through the same channels of perception and objectification. Nothing in the specialized corpora of these agents maps to the operations of the systems that supposedly create them, and hence these operations are inexpressible in terms of those very systems. What is needed is an account of not just knowledge but of the nature of this knowledge and the innate conditions of acquiring it.

Kant's philosophy provides such an account of the nature and acquisition of knowledge. His categories provide a general-purpose ontology for knowledge that maps onto the core operations of the common model. Kant aims to explain how we construct appearances so that our experience of the world is not a rhapsody of perception, but a unified experience with cognitively constructed objects. However, most cognitive models and cognitive architectures take objects as given rather than constructed. Given that the goals of both the common model enterprise and Kant are to be psychological (rather than to merely create agents optimized for a specialized task), the same needs to be reflected in the choice of memory architecture. For this reason, we look at vector symbolic architectures.

Vector Symbolic Architectures (VSAs) are a family of connectionist architectures specifically developed to create systems for symbol instantiation and manipulation (for a detailed review refer to [11]). VSAs operate via algebraic operations on distributed representations over a high-dimensional vector space. The exact choice of vector operations varies from model to model, but what is common across VSAs is that things are represented by a vector defined by a set of randomly generated numbers (e.g., [-0.05, 0.72, -1.72, 0.14…]). Relationships are stored by moving the corresponding vectors through a single high-dimensional space in which they are instantiated. VSA-based models have gained popularity due to their neurally plausible architecture, and ability to model psycho-linguistic phenomena including semantic priming, similarity, and association [12],[13] as well as basic memory phenomena. For example, the DSHM model has been used to demonstrate the fan effect—a delay in the recall of things with a larger number of irrelevant associations [14].

For Kant, our knowledge of the world stands united in our mind and his categories "play a role in constituting objects from the sensory material, so that it seems wrong to take objects as given from the outset" [10]. Given the ability of VSAs to represent concepts in a unified high-dimensional geometrical space and the fact that it "aims at modeling 'concepts in the head' rather than 'things in the world'"[15], VSAs stand out as a particularly good choice for memory system that deploys Kantian categories. Doing so also addresses a critical concern raised in the VSA literature regarding a lack of logical structure to the relationships between concepts represented in these systems. This issue, called the bag-of-words/bag-of-concepts problem, is that VSAs "…indicated a degree of similarity between two items, [but] not any particular relationship since the vectors are inherently 'non-meaningful'. The knowledge representations make no commitment to a particular set of features or theory of meaning, although the vector

‡ Critique of Pure Reason, A 76/B 101

representations imply a certain degree of relatedness in order to model cognitive effects”[16]. This bag-of-words issue can be resolved by deploying the categories proposed by Kant in a VSA model. Thus, just as Kantian categories can benefit from being instantiated as VSAs, VSAs can benefit from the structure provided by Kantian categories.

A VSA-based system called DSHM [17]deviates from other VSA models in that the modelers add innate structure to memory. By introducing cardinal vectors as innate atomic items, Rutledge-Taylor et al. make a distinctly Kantian move. While for DSHM, the cardinal vectors were something created ad hoc as features for models of specific tasks, we create a system called Kantian Holographic Declarative Memory (K-HDM) [18] in the programming language R that makes cardinal vectors a core element of the model. The model is available on GitHub. Cardinal categories in K-HDM represent the Kantian categories and are consistently used to encode all conceptual relationships. This satisfies Kant’s conceptualization of categories as universal and necessary.

5. **Kantian Holographic Declarative Memory**

All relations in K-HDM are stored as subject–relation–object predicates written as ***xRy***, where *x* is the subject, *y* is the object and *R* is the relation described via the Kantian categories. *R* thus becomes the representative of the Kantian categorical ontology. To take an example, traditional models would treat the following three ***xRy*** relationships—*Sun heats earth, Dogs make noise, Love conquers all*—as entirely distinct. Yet, understood through Kantian *categories*, the subject and the predicate in all three are related by the same Kantian category types (see Table 2). Any relationship can be thus described by the selection of one type from each of the four *categories,* giving us a total of 3^4=81 basic relationships. For examples, see Table 3.

Table 2. Kantian Categorization

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

Table 3. Examples of Categories for Different Relations

Relationship	Quantity	Quality	Relation	Modality
Bachelors are unmarried (he definition of *bachelors*)	Universal	Affirmative	Categorical	Apodeictic
Flu is either viral or bacterial	Universal	Affirmative	Disjunctive	Assertoric
Bill makes cake	Singular	Affirmative	Hypothetical	Assertoric
Mary is not a doctor	Singular	Negative	Categorical	Assertoric
Relationship	Quantity	Quality	Relation	Modality

5.1. Encoding in K-HDM

Like DSHM, K-HDM also represents concepts through two vectors—environmental and memory. An environmental vector ($\boldsymbol{E}_c$) represents the percept of the concept (**C**), and can be understood as its referent. A memory vector ($\boldsymbol{M}_c$) encodes **C**’s 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. Given this encoding schema, information such as “*some doors are red”* will have a memory chunk as shown in Table 4.

Table 4: 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_{pat}	–
Quality	Affirmative	E_{aff}	–
Relation	Categorical	E_{cat}	–
Modality	Assertoric	E_{asr}	–

The relation *R*, which implements the Kantian categorical ontology, is represented by *Relationship vector* (denoted by $\mathbf{Rel_{vec}}$). This vector is created by convolving together the cardinal vectors of the relevant category types. Using this vector, the information about the subject and the predicate is encoded in the memory vectors. For example, for information like *some doors are red* we create the relationship vector (see equation 1) and update the memory vectors for *red* and *door* (see equation 2 and 3). We use non-commutative circular convolution (denoted by *; see [19]) to create conjunctive relations between concepts and vector addition (denoted by +) to add these relations to their meanings vectors ($\boldsymbol{M}_c$).

$$Rel_{vec} = E_{pat} * E_{cat} * E_{aff} * E_{asr} \quad (1)$$
$$M_{door} = Old\ M_{door} + (Rel_{vec} * E_{red}) \quad (2)$$
$$M_{red} = Old\ M_{red} + (Rel_{vec} * E_{door}) \quad (3)$$

5.2. New Concepts and Complex Relations in K-HDM

Unlike DSHM, K-HDM stores all *x*R*y* relationships as new concepts with their own pairs of 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 values of the vectors of the concept are created by convolving the environmental vectors of all constituent concepts and categories (see equation 4). This new concept can then be used to create more complex concepts through the same process by which it was encoded. Complicated relations such as *the sun heats the earth* can be encoded through Kant's *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*.

$$E_{P1} = M_{P1} = E_{door} * Rel_{vec} * E_{red} \quad (4)$$

5.3. Results from Simulations

Simulations run on Kantian-HDM replicate psychological behavior (like the fan-effect [14],[17]) as well some other interesting effects courtesy the addition of Kantian categories. For example, K-HDM allows for the analogical reasoning of the kind 'A is B as C is to ?'. Widdows and Cohen [20] deploy something akin to the hypothetical Kantian category in their vector system to get similar results for analogical reasoning. In what follows, we will focus on the unique results obtained in K-HDM regrading quantitative estimates.

K-HDM has multiple control parameters one of which is 'Repeat Storage'. This parameter allows for control over whether repeated exposure to the same proposition, e.g., 'Ball is Broken', results in repeated storage of the relationship. Turning on the feature results in an interesting property where the model does not require explicit storage of some of the Quantity category types (Universal and Particular) but can rather derive the types from the 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 the shadow that a vector would cast on another vector (Figure 3). The projection of vector A on vector B can be computed by multiplying the magnitude of

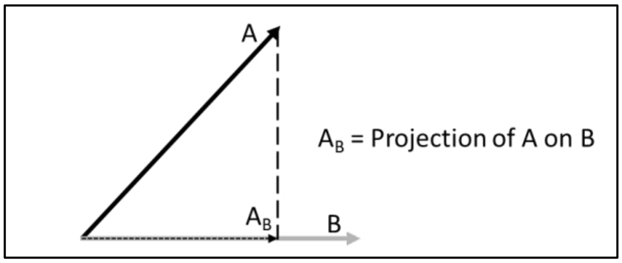

Figure 3: Dot Product as Projection of a vector onto another

A with the cosine of the angle between A and B.

To demonstrate this, we generated 20x10 simulated datasets. In each, 10 balls are randomly tagged either *broken* or *not broken*. The model stores *Ball is Broken* for every broken ball and *Ball is not broken* for every unbroken ball. At the end of storage, a ratio of the projections (dot products) of the memory vector of *Ball* on the vector representing *is broken* and *is not broken* allows the model to estimate the actual ratio (see equation 5). This ratio is used to derive the quantity relationships. Figure 4 shows results averaged over 20 simulations and the error margin.

$$Esitmated\ Ratio\ of\ Borken\ balls\ to\ Unbroken = \frac{Projection(M_{Ball}, M_{is\ broken})}{Projection(M_{Ball}, M_{is\ NOT\ broken})} \quad (5)$$

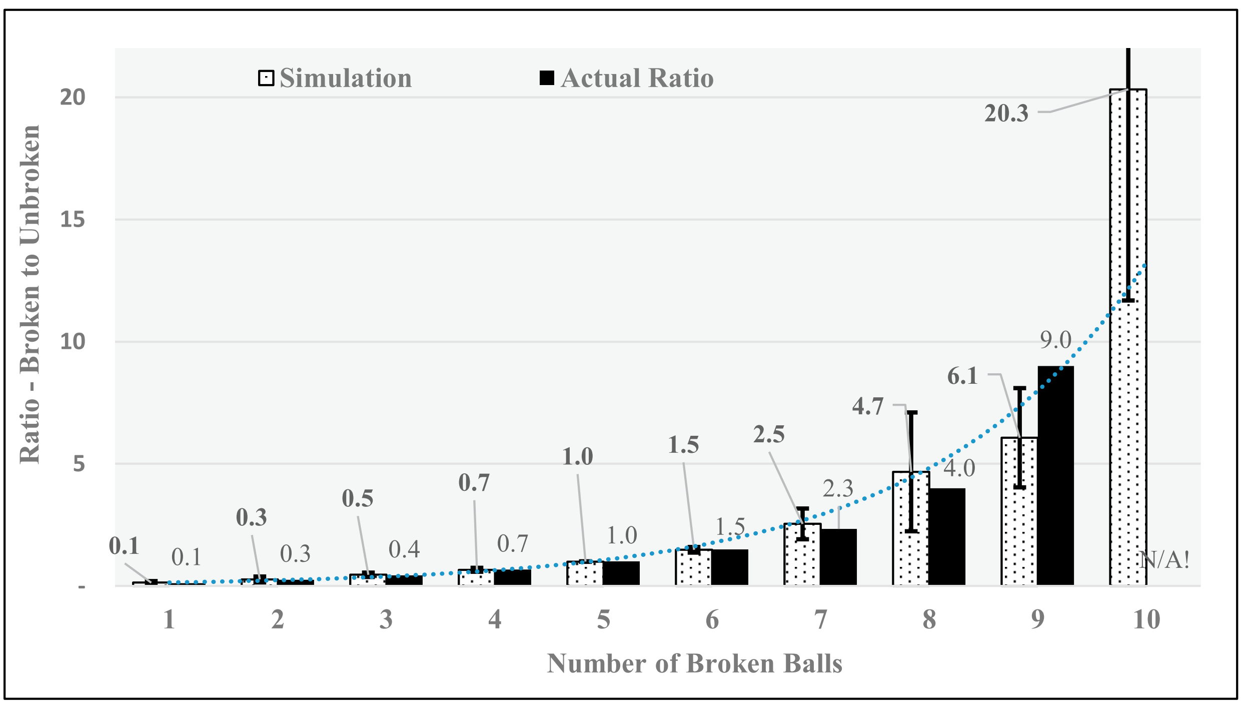

Figure 4: Projection Ratio and Quantity estimates

The estimated ratio grows in a pattern similar to the actual ratio of *not broken* to *broken*. The correlation between the actual ratio and estimated ratio is high ($R = 0.96, p \leq .001$). The trendline in the graph demonstrates the exponential nature of the result. Something to note here 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 a divide by zero. The estimated ratio shows a huge jump for this condition. Such jumps are sufficient flags for the model to determine an extreme condition like the *Universals* (*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. What we see here is the evidence of compatibility between Kantian categories and the vector architecture. A similar claim is made in [20] concerning the ability of VSAs to express negation (which is also a Kantian category) in their work. Our future work will be aimed at expressing the other categories using features of the vector symbolic algebra.

6. Conclusion

Although separated by more than two centuries, Kant's *Critique of Pure Reason* and the common model enterprise share much in common. The two are united by the goal to isolate that foundational architecture of the mind that forms the essence of human-like cognition. The emphasis Kant's philosophy puts on the integrated unity of the mind makes it especially relevant for the common model enterprise which is primarily an exercise of integration. Kant wrote before the cognitive revolution and thus did not have the advantage of its computational language. Yet his thought was far ahead of his time. His normative philosophy can provide a glimpse into future questions that the common model ought to answer in order to achieve its goal. On the other hand, cognitive modeling can provide a much-needed clarification and an exact implementation of Kant's work.

A question relevant to the common model enterprise was raised at the beginning of this paper—*what are the conditions that constitute the ground for bounded rational behavior?* Kant finds an answer to this in the objective nature of our knowledge and experience which is created through the interaction of the faculties of mind. In doing so, Kant elevates the importance of something akin to Newell's knowledge level [3]. It describes the relationship between memory systems that store knowledge and the cognitive operations that manage the production and use of this knowledge. One of the ways to realize this knowledge level is through the implementation of the Kantian categories in the memory system. Because the categories are reflected in the operations of the cognitive architecture on one hand and the pure concepts present in our knowledge, on the other hand, they bridge the gap between the contents of memory and the cognitive systems that supposedly create them. We create a vector-based memory system called K-HDM that implements Kant's categories, and further demonstrate its function through simulations.

Acknowledgments

This research has been funded by an NSERC PDF and NSF grant BCS-1734304 to M. A. Kelly.

References

[1] Laird JE, Lebiere C, and Rosenbloom PS. (2017) "A Standard Model of the Mind: Toward a Common Computational Framework across Artificial Intelligence, Cognitive Science, Neuroscience, and Robotics." *AI Magazine* **38** (**4**): 13+.

[2] Kant I. (1781) "Critique of Pure Reason." Hackett Publishing Company.

[3] Newell A. (1981) "The knowledge level." *Artificial Intelligence* **18** (**1**): 87–127.

[4] Russell B. (1948) "Human knowledge, its scope and limits." Simon and Schuster.

[5] Guarino N. (1995) "Formal ontology, conceptual analysis and knowledge representation." *International Journal of Human-Computer Studies* **43** (**5**–**6**): 625–40.

[6] Clancey WJ. (1989) "The knowledge level reinterpreted: Modeling how systems interact." *Knowledge Acquisition: Selected Research and Commentary*, Boston, MA, Springer, 39–45.

[7] Williams G. (2018) "Kant's Account of Reason." in Zalta EN (ed) *The Stanford Encyclopedia of Philosophy*, Metaphysics Research Lab, Stanford University.

[8] Clark A. (2013) “Whatever next? Predictive brains, situated agents, and the future of cognitive science.” *Behavioral and Brain Sciences* **36** (**3**): 181–204.

[9] Neurath O. (1983) “An International Encyclopedia of Unified Science.” in Cohen RS, Neurath M (eds) *Philosophical Papers (1913–1946)*, vol. 6, Dordrecht, Springer, 139–44.

[10] Achourioti T, and van Lambalgen M. (2011) “A formalization of Kant’s transcendental logic.” *Review of Symbolic Logic* **4** (**2**): 254–89.

[11] Mitchell J, and Lapata M. (2008) “Vector-based Models of Semantic Composition.” *Acl* **8** (**June**): 236–44.

[12] Jones M, Kintsch W, and Mewhort DJK. (2006) “High-dimensional semantic space accounts of priming.” *Journal of Memory and Language* **55**: 534–52.

[13] Jones M, and Mewhort DJK. (2007) “Representing word meaning and order information in a composite holographic lexicon.” *Psychological Review* **114** (**1**): 1–37.

[14] Kelly MA, Kwok K, and West RL. (2015) “Holographic Declarative Memory and the Fan Effect: A Test Case for A New Memory Module for ACT-R.” *International Conference on Cognitive Modeling (ICCM)*, Netherlands, , 148–53.

[15] Kornai A, and Kracht M. (2015) “Lexical Semantics and Model Theory: Together at Last?” *Proceedings of the 14th Meeting on the Mathematics of Language (MoL 2015)*, Stroudsburg, PA, USA, Association for Computational Linguistics, 51–61.

[16] Burgess C, and Lund K. (2000) “The dynamics of meaning in memory.” *Cognitive Dynamics: Conceptual and Representational Change in Humans and Machines* **13** (**May**): 17–56.

[17] Rutledge-Taylor MF, Kelly MA, West RL, and Pyke AA. (2014) “Dynamically structured holographic memory.” *Biologically Inspired Cognitive Architectures* **9**: 9–32.

[18] Arora N, and West R. (2018) “Kantian Holographic Declarative Memory Module in R.”

[19] Plate TA. (1995) “Holographic reduced representations.” *IEEE Transactions on Neural Networks* **6**: 623–41.

[20] Widdows D, Cohen T, Bing M, and Cohen T. (2014) “Reasoning with vectors: A continuous model for fast robust inference.” *Logic Journal of the IGPL* **23** (**2**): 141–73.