

Exploring the Reality of the Knowledge Level: Pragmatism Embodied

**By
Jeremy T. Young**

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

**Master
in
Cognitive Science**

**Carleton University
Ottawa, Ontario**

**©2018
Jeremy T. Young**

Abstract:

Allen Newell's Knowledge Level (KL) theory is a philosophical position on the reality of knowledge explained by Pragmatic principles—most notably that the practical effects of general concepts justify their reality. The KL's existence is context-dependent and task-specific based on its practical effect—goal attainment. It reduces the complexity of behavior prediction by abstracting away irrelevant details. The behavior patterns identified by the mechanism are real because they endure, are observable, are and interpretable across individuals. Newell's computational models describe the brain as a Pragmatic engine. The KL theory is similar to previous theories by Karl Popper and Daniel Dennett.

Viewing the KL as real enhances precision of testing cognitive theories by computationally modeling behavior and highlights the dynamic role values play in re-prioritizing goals as knowledge is acquired. Newell's research could be expanded to model how manipulation of agent's knowledge/goals enables influence and deception, for use in predictive analytics.

KEYWORDS: Knowledge Leve l; Knowledge Representation; Cognitive Modeling; Cognitive Architectures; Inter-theoretic Levels; Agency; Behavioral Prediction.

Acknowledgements:

I would like to express deep appreciation for my graduate advisor Dr. Robert West and my graduate mentor Dr. Andrew Brook. These two helped me maintain a balance of pushing the envelope on a difficult issue while ensuring I recognized the full implications of the paper's statements. Dr. West's cognitive modeling course initially inspired me to apply my philosophy background to explore the philosophical underpinnings of this issue and its application to cognitive science and artificial intelligence (AI). My regular dialogues with Dr. West on cognitive modeling and rational actor models helped me recognize the crucial need for more research into how values guide goal selection, deconfliction, and prioritization in agents. Dr. Brook regularly made himself available for in-depth discussions on the broader areas of the age-old ontology-epistemology debate—keeping me focused on the aspects of the debate relevant to cognitive science and AI. Dr. Brook was particularly helpful in giving his project the final push it needed to see completion. I would also like to thank Deniz Askin for being my counterpoint as I explored ideas related to the thesis.

Most importantly, I would like to thank my wife Maria Zuniga—her unending support enabled me to balance a graduate degree with my career and our beautiful daughter. She was always there to inject motivation when I was flagging from fatigue and, ironically, to ensure that this project remained focused on practical applications.

Table of Contents:

TITLE PAGE	i
ABSTRACT	ii
ACKNOWLEDGEMENTS	iii
TABLE OF CONTENTS	iv
LIST OF TABLES	vi
INTRODUCTION	1
PART I: Reality, Truth, & Knowledge in Newell's Theory (What does it mean to be real)	5
1.0 What is Real?	6
1.1 Neural Realism	9
1.2 Pattern Realism	10
1.3 Knowledge as Meaning	11
PART II: Understanding the Knowledge Level	13
2.0 The Use of Levels in Cognitive Science:	15
2.1 Marr Levels vs. Newell's	16
2.2 Epistemological Use of Levels	17
2.3 Ontological Use of Levels	20
2.4 Hybridized Use of Levels	23
2.5 Context and Boundary Conditions	25
3.0 The Cognitive Level:	26
3.1 Knowledge is Not Found within the Cognitive Level	27
3.2 Cognitive Mechanisms Exist when Boundary Conditions are Met	28
3.3 Cognitive Modeling Architectures are Task-bound Simplifications	29
4.0 The Knowledge Level:	30
4.1 Components of the Knowledge Level	31
4.2 The Knowledge Defines Boundary Conditions for Problem Spaces	33
4.3 Knowledge as the Medium	34
4.4 Impetus Behind the Genesis of the Knowledge Level Theory	36
PART III: The Knowledge Level as Pragmatism Embodied-Situating the KL in Pragmatism	37
5.0 The Pragmatic Framework of the Knowledge Level:	39
5.1 The Origins of Pragmatism	39
5.2 What is Realism for Pragmatism?	40
5.3 A Quick Note on Instrumentalism vs. Pragmatism	41
5.4 The Real is What Furthers our Goals	42
5.5 Knowledge Transforms the Physical World	45
5.6 Knowledge is Acquired by Doing	45
5.7 Everything Experienced Must Somewhere be Real	46
5.8 Application of Pragmatism to the Theory of Knowledge	47
5.9 The Knowledge Level in the Scientific Method	48
5.10 Context Determines Truth	49

6.0 Pragmatism's Grounding in Empiricism:	51
6.1 The Knowledge Level Bridges Empiricism and Rationalism	51
7.0 Newell's Model Instantiates Pragmatism:	53
7.1 Newell's Model of the Brain as a Pragmatic Engine	53
7.2 Computational Expressions of Pragmatism's Principles	54
PART IV: The Knowledge Level Theory as a more Committed Perspective	56
8.0 Why does the Ontological Nature of the Level Knowledge Level Matter?	57
9.0 Popper's Rationality Principle:	58
9.1 Popper's Criteria for Determining the Scientific Nature of Theories	59
9.2 Contradictions with Popper's Rationality Principle and Falsification	60
9.3 Popper's Process of Scientific Discovery	61
9.4 Popper's of Scientific Discovery	63
9.5 The Rationality Principle's Pragmatic Goal-Directed Approach	65
9.6 Distinction between Popper's Theory and Newell's	66
10.0 Dennett's Intentional Stance:	67
10.1 The Role of Intentionality	68
10.2 The Context-dependent Nature of the Intentional Stance	68
10.3 Dennett's Perspective on Levels	72
10.4 Where does Dennett's Theory Fall on the Epistemic-Ontic Scale?	73
10.5 Similarities between the Intentional Stance and the Knowledge Level Theories	77
10.6 Similarities between Dennett's Views and Pattern Realism	79
Conclusion:	80
References:	83

List of Tables:

TABLE 1	Marr/Dawson's epistemic levels system	19
TABLE 2	Example of the element comprising 2 of the systems levels	22
TABLE 3	Mediums of computer systems levels	22
TABLE 4	Newell's use of levels and bands	24
TABLE 5	Medium and law pairings for levels	31
TABLE 6	Knowledge level elements correlated to computer systems level elements	34
TABLE 7	Mapping of terminology between Newell and Dennett's theories	78

Introduction:

Despite being one of the most influential figures in cognitive science and artificial intelligence, Allen Newell's most radical idea—that the knowledge level was real—has been largely overlooked by researchers coming after him. Although theoretical research has been done on Newell's Knowledge Level theory (see Lieto, Lebiere, & Oltramari, 2017 for a recent example), the reality claim has not received the broader attention it deserves. According to Newell, "Distinguishing this level as real leads to a simple and satisfactory view of knowledge and representation...it dissolves some of the difficulties and confusions that we have about AI" (Newell, 1982, p.13). Although Newell's theory focuses on arguing for the ontological reality of knowledge, it is important to point out that the manipulation of knowledge in the service of problem solving is only one of several key tasks that happen at the upper level of cognition—commonly referred to as the 'computational task' or 'procedural' level of cognition. Others such as memory, internal rotation of 3D objects, and issues of perception are also higher-order cognitive activities worthy of discussion, however, they will not be topics of this paper.

Ontologies describe the set of what exists in the world (i.e., what the actual structure of objective reality is), whereas epistemologies are the frameworks we employ to analyze and ask questions regarding the subset of data (knowledge) about the world that reaches us through our senses. There is a difference between the real ontologies in the objective world and ontological beliefs which exist in our heads—which may or may not be true/real—about what objects and phenomena exist in the world. I will call

external objective reality *the real ontology* and I will call beliefs in our head ontological *beliefs or claims*.

It is important to make this distinction because the term ‘ontology’ is often loosely employed to mean both what *the reality* existing outside of the human mind is and what people *believe* that reality to be and how these beliefs guide their behavior. Another confusing factor is that many epistemologies contain ontological beliefs. In these cases, we can say the epistemology involves acting “as if” a set of ontological beliefs are true (James, 1929). For example, the strict separation between ontological beliefs and an objective ontological reality that has just been defined is an epistemology, but it is also an ontological belief. Scientists act “as if” this ontological belief is true (epistemology) and most believe that it is true (ontological belief). A person who believes in witchcraft would likely not use the epistemology of science and would reject its corresponding ontological beliefs. Examples of using an epistemology that contains an ontological belief, without accepting the ontological belief as true, include work on the Higgs Boson (before it was found to actually exist), as well as early interpretations of Einstein’s theory of relativity and psychologists who view their theories as ‘useful fictions’.

Newell is not an ‘as if’ theorist—he thought there was something real/actual about the mechanism that allows us to interpret and predict the behavior of other agents and plan our own behavior to maximize attainment of our goals. When Newell said that, ‘the knowledge level is real’, he was making an epistemological claim that

contains an ontological belief. The ontological belief is that the knowledge level is ‘real’. His epistemological claim is that acting “as if” this belief is true produces advantages over epistemologies that assume this belief is false—i.e., that it provides explanatory grip. Newell clearly believed that the knowledge level was real, but he sometimes stated the matter in terms of the advantages of acting “as if” the knowledge level is real, showing that he was aware of this distinction as he attempted to communicate his ideas to others.

Newell’s realism is a rather uncommon variety. Looking at him superficially, it is difficult to understand why he positions himself as a realist. It is also not clear whether he would have declared himself to be a Pragmatist, however, his work exhibits a very Pragmatic character and this Knowledge Level theory finds support in several principles of early Pragmatism.

Newell was not a philosopher and did not explicitly develop the philosophical underpinnings of his position. To understand Newell’s work, it is essential to recognize the similarities between his ideas and the foundational works of Pragmatism. By Pragmatism, I mean the work of its original authors such as Charles Sander Peirce, William James, and Ferdinand Canning Scott Schiller, and not the myriad—and often highly divergent—views calling themselves Pragmatism that came afterwards. The thesis explored in this document is that Newell’s writing and his contributions to AI can be more clearly understood through the lens of Pragmatism, that they represent an articulation and extension of *Pragmatism*, and that viewing the cognitive mechanism he

describes enhances the precision of our testing of theories of cognition through models and simulations.

I wish to apologize in advance if in the process of trying to elucidate and further develop Newell's work, I have inadvertently mischaracterized his original intent. I will defer to those who knew Newell personally and were more directly involved in his research and theorizing to pass final judgment in this area. What is presented here is an interpretation of Newell's philosophy which seems right as it does, as he claimed, dissolve difficulties and result in a simple and satisfactory view of knowledge.

Roadmap for the Paper

- ❖ **Part I: Reality, Truth, & Knowledge in Newell's Theory (What does it mean to be 'real'?)** explains that the answer to this question depends on one's particular brand of realism, variations of which are prolific and highly divergent in the fields of philosophy, psychology, and cognitive science. Critics to theories such as Newell's often cite concerns regarding *correspondence* (between ideas of the mind and what lies outside the mind) and *foundationalism* (what is at the base of all reality) as road blocks to viewing cognitive phenomena such as knowledge as being real entities. Newell was not concerned with determining ultimate objective truth, rather he was focused on discovering the role that belief played in what agents think is true, which subsequently guides their actions. That is, he wanted to understand how knowledge of our world shapes our behavior. Newell's version of what is real was context-dependent and guided largely by the reliably predictable patterns of meaning that can practically be shown to provide insight into the actions of others. Because these patterns drive agent actions that continually reshape the physical world, they must themselves have some sense of reality.
- ❖ **Part II: Understanding the Knowledge Level** explains Newell's thoughts about the knowledge level and the role that knowledge and intentions play in human cognition and prediction of behavior. It highlights how Newell's theory provides a framework for reducing the complexity of analyzing every mechanism behind intelligence systems by abstracting away details irrelevant to the prediction of intelligent behavior. As such, the theory is critical in light of current challenges involving agency and knowledge representation in cognitive science and artificial intelligence research.

- ❖ **Part III: The Knowledge Level as Pragmatism Embodied (the relation between the KL and Pragmatism)** elucidates the links between Newell's knowledge level and the original principles of Pragmatism. Pragmatism focuses on the practical, observable impact of our knowledge/beliefs, and views meaning and truth in terms of the effects that believing an idea would have in practice—views which support the reality of the knowledge level. Newell's work, in turn, embodies the Pragmatic principles in computational models that reveal how the brain can be viewed as a Pragmatic engine. Without directly referring to itself as Pragmatism, Newell's theory is more philosophically grounded in the original principles of Pragmatism than most discussions in cognitive science and AI research which employing that term.
- ❖ **Part IV: The Knowledge Level Theory as a more Committed Perspective** explains that Newell was not the first theorist to describe the cognitive mechanism that allows us to use knowledge and goals to predict the behaviors of other agents. Karl Popper and Daniel Dennett described similar theories but stopped short of claiming this mechanism was real. By including the ontological belief that the knowledge level is real within his epistemology, Newell avoided certain pitfalls inherent in using the knowledge level as an epistemology while denying it is reality—issues both Popper and Dennett ran into—and he laid the foundation for using cognitive models to enhance the precision of our evaluation of these theories.

PART I: REALITY, TRUTH, AND KNOWLEDGE IN NEWELL'S THEORY:

What does it mean to say that “the knowledge level is real”? The answer to this question depends on one’s variant of realism. Variations on the term ‘realism’ abound in philosophy, psychology, and cognitive science. A common objection to Newell’s assertion that the knowledge level is real is that it seems at odds with a scientific worldview. Many people seem to have a gut reaction that what Newell is talking about is not real, but there is some evidence for considering that it might be. Newell’s version of realism was an uncommon one, but one worth evaluating in a deeper light.

There appear to be two primary concerns or doubts about the reality of phenomena like the knowledge level. The first concern that many people have turns on the issue of *correspondence*. This is the view that there is a reality outside of the human mind (an ontology) and our understanding of that reality (epistemology) needs to correspond to it. This is a type of scientific realism. Scientific realism is currently very popular. If one subscribes to a pure strain of this view, then they would likely reject both Newell and Pragmatism's perspective on reality.

The second concern that people who take issue with Newell's stance on the reality of the knowledge level have deals with the idea of *foundationalism*. This is the view that reality is built upon some foundation and this foundation must be real. For cognition, this foundational or implementational level is broadly considered to be neurons, with higher order cognitive activities arising from various neural assemblies to produce a range of functions that support intelligent action. This view is a more generalized version of neural realism.

1.0 WHAT IS REAL?

Most of us believe there are real things in the world. The predominant perspective in modern science is based on a fundamental assumption that a correspondence between knowledge and objective reality is possible. Indeed, this is what justifies the significant expenditure of effort toward scientific endeavors. The descriptive work of science results in collective ontological claims—i.e., claims about what exists and the descriptions of the phenomena in the world as humans understand

them. These ontological claims are theories about what is and is not part of the ontology (objective reality).

Scientific realism is a perspective on the division between ontological beliefs and ontological reality. Although broad variations exist in the use of this term, scientific realism is defined by the Stanford Encyclopedia of Philosophy (SEP) as “a positive epistemic attitude toward the content of our best theories and models, recommending belief in both observable and unobservable aspects of the world described by the sciences” (Chakravartty, 2017). In other words, we should accept ontological claims that pass a certain threshold for scientific realism, as corresponding with the actual ontology, until such time as science disproves them.

Post-modernism is a different perspective on ontological claims that denies an objective reality and/or our ability to come to know this sort of reality. As with scientific realism, there is a range of variation, however, the central theme of most definitions of post modernism is an opposition to universal notions of objective reality, truth, morality, and reason (Aylesworth, 2015).

Pragmatism is a perspective that lies somewhere in between these two views. Pragmatism defines the practical part of a discussion of reality—the only part about which it concerns itself—as being defined by how a belief (or ontological claim) is true/real with respect to a given context (environmental situation) and goal. It focuses on how this reality is justified by its measurable effects/impact. For Pragmatism, the

real is that which has impact on our ability to solve problems. Pragmatism overlaps with scientific realism because knowledge that leads to correct predictions is useful and therefore “true” in the pragmatist sense. Unlike post modernism, Pragmatism and scientific realism can use prediction and falsification to evaluate truth. However, Pragmatism also overlaps with a certain sense of post modernism’s sense of relativity because truth is relative to the context at hand. A piece of knowledge can be true in one context and false in another. This violates scientific realism, which seeks to establish certain ontological beliefs as true, regardless of context. However, the disagreement is not about the existence of an objective reality, the disagreement is about what constitutes truth. For scientific realists, something is true if it can be reliably replicated under specific conditions. For Pragmatists, something is true if it takes you closer to your goal. In other words, scientific realism is concerned with *what is* and Pragmatism is concerned with *what you ought to do*, given a particular goal.

Newell’s concern was not with identifying some ultimate objective truth, it was with describing how belief (true or false) guides the actions of intelligent agents. Because we are creatures produced through evolution, our concern is always, ultimately, about what we ought to do to survive. Although Newell did not extensively discuss the connection to evolution, he believed his system was consistent with an evolutionary framework—in particular Simon’s (1969) work on evolution (for more recent discussions of how our use of knowledge is a factor of our evolutionary struggle, see Jordan Peterson, 1999). It can be argued that Pragmatism is more appropriate than scientific realism for understanding human cognition because it focuses on the issues

relevant to the problem-solving required for our species to survive and flourish. If we accept that humans are fundamentally pragmatic, then scientific realism can be subsumed under pragmatism as a belief that can play an important role in increasing the amount of knowledge available to support our decision-making (i.e., by providing stable sets of facts and laws to aid in attaining our goals).

1.1 Neural Realism

Neural realism is a sub category of the more general and intuitive view of science that is often called pure physicalism or physical realism—the view that only those things that possess clearly sensible physical qualities should be considered real. Neural realism is the view that the mind and all its intentional constructs/functions map back (i.e., materially correspond) to physical networks of interconnected neurons in the brain. Neural realism is currently the predominant view in the fields of psychology, cognitive science, and neuroscience. Under a pure interpretation of this view, the existence of symbols in the brain is false, unless by the term “symbol” one means patterns of connections between neurons and the individual weights of those neurons (i.e., a specific sub-network(s) in a connectionist architecture). This view interprets the classical theory of the mind as an epistemology—a convenient fiction, schema or epistemological scaffolding that serves no other purpose than to help us make sense of what is in all reality a more complex, though completely physical, process. The physical structure of the brain makes the argument for neural realism compelling.

In this view, only neurons and neural structures are “real” and ontological claims (theories) about cognition can only be made for neurons and neural structures. Everything else is typically viewed as an epistemology for understanding how neurons work—that is, they are convenient fictions. Neural realism gives neurons a privileged status. Dennett seems to flirt with Neural Realism because he appears to believe that there is such thing as memory even though we do not know how neurons make this cognitive phenomenon happen. Yet his attitude towards intentions is somehow treated differently. This is because he wants to wait to commit himself to attributing reality to the intentional stance until neuroscience can map the physical structures of the brain in such a way they can explain the proven functionality in the intentional stance and similar theories.

1.2 Pattern Realism

There appears to be strong agreement within the cognitive science community that the physical brain is constructed of neurons organized into large networks (neural assemblies). Where the disagreement arises is whether the brain also contains higher level constructs such as symbols, functions, logic, knowledge, etc. Specifically, there is disagreement over whether these higher-level constructs should be considered epistemologies or part of the actual ontology of the world; that is, are they real in the sense that neurons are real?

According to neural realism, everything is up to neural circuitry and after that all the higher-level processes explained by symbols are fantasy that we use to help

organize our understanding of what brains do. Within this view, the knowledge level cannot be real. However, there is another way to view reality, a view which has been articulated by a number of different theorists (e.g. Newell, 1994; Chomsky, 2016; Hofstadter, 1999; Peterson, 1999, Dennett, 1987). There is no agreed-upon name for this view so for clarity I will call it, *Pattern Realism*. In this view, reality is equivalent to the patterns that can be stably instantiated across a range of different substrates/media. Only the quantum foundations are objectively real in a physical sense because quantum physics is the foundation, out of which the field of chemistry emerges as a stable, predictable set of patterns. Similarly, biochemistry emerges from chemistry, and neuroscience emerges from biochemistry. Viewed from the perspective of physics, neurons are stable biochemical patterns and do not have a privileged status. Their objective existence is a factor of their stability and predictability. Informational structures such as symbols, knowledge, beliefs, goals, etc., are patterns that also display stability and predictability and therefore, according to pattern realism, would have the same ontological status as neurons, that is, they are real.

1.3 Knowledge as Meaningful Pattern

With the above in mind, I argue that knowledge is real because its data structures exist as stable patterns that are always physically instantiated, exist within specific boundary conditions/context, and cohere across time. Because knowledge is a pattern, it requires both *form* (a relation among parts) and *content* (the material in which the pattern is instantiated)—you cannot have one without the other. The value of entities typically resides in their relations, because relation between parts

predominantly determines function (in the case at hand, function is predictability). For example, a chair may be made of steel, wood, or plastic but it serves the same function based on the relation that exists between its various parts. There may be some variation in this relation, but the core relation remains, or the function suffers.

Such a definition accounts for the manner in which knowledge leaps from one individual to another. That is, knowledge is transferred from one agent to another via representations (patterns) as long as the two agents share the same high-level language. The patterns that are transferred can be exchanged through a broad range of physical instantiations (substrates) including: sound waves, magnetic patterns, printed type, digital pixels, etc. and they are stored in the agents as patterns in neural assemblies. The fact that one agent can have a thought (knowledge) and that the thought continues to exist when the first mind that originally conceived it dies (i.e., the neural assembly that initially held it no longer exists) further indicates that the knowledge is not dependent only upon neurons. That is, the knowledge is a stable pattern that endures across a range of physical substrates (neural networks, print, digital media, etc.) as long as the boundary conditions of the pattern endure.

Newell's version of realism is a belief in the reality of the knowledge level, but one that recognizes that the neuron has traditionally been given a privileged status (i.e., it is commonly thought of as being 'more real' by many in the cognitive science community—particularly among those who specialize in neuroscience). The neural level is not more or less real, we simply know more about it. Though while we know many

things about its description (i.e., its physical qualities), we know little about how the activity of neurons initiates or guides behavior in agents.

Part II: Understanding the Knowledge Level

According to Newell, “To treat a system at the knowledge level is to treat it as having some knowledge, some goals, and believing it will do whatever is within its power to attain its goals, in so far as its knowledge indicates” (Newell, 1982, p.13). Newell used the term, *Rationality Principle*, to describe the operating mechanism behind the knowledge level. According to Newell, the rationality principle states that an agent will use the knowledge it has of its environment to achieve its goals (Newell, 1982, p.17). Newell later elaborated, “a person acts to as to attain his goals through rational action, given the structure of the task [context] and his inputs of information and bounded by limitations on his knowledge and processing ability” (Newell, 1994, p.33). Although Newell does not appear to have explicitly referenced Popper, he was likely referring to Popper’s Rationality Principle as his description seems identical and given Newell comments elsewhere on Popper’s perspective on science (Newell, 1994, p. 14). Popper’s version of the Rationality Principle was the narrowest sort of instrumentalist rationality—given an agent’s aims (goals) and knowledge (beliefs), we can assume “that the various persons or agents involved will act *adequately*, or *appropriately*—that is, in accordance with the situation (Popper, 1994, p.169). Similarly, Dennett explained that in the intentional stance, “You decide to treat the object whose behavior is to be

predicted as a rational agent; then you figure out what beliefs that agent ought to have, given its place in the world and its purpose...then you figure out what desires it ought to have...and finally you predict that this rational agent will act to further its goals in the light of its beliefs (Dennett, 1987, p.17).

All three theories discuss the mechanism of the human mind that translates descriptions of an agent's environment (scientific knowledge) into directives for action (intentions) and explain how we can predict agent behavior in light of this. Although there is much disagreement in the larger scientific and philosophic community regarding what constitutes true 'knowledge' and what separates it from beliefs (which may or may not be false), I will employ the term 'knowledge' in the way Newell used it, which can include beliefs that are true or false as long as the agent's behavior is guided by those beliefs. The simple equation for Newell, Dennett's and Popper's theories effectively amounts to:

{GOALS + KNOWLEDGE/BELIEFS + RATIONALITY PRINCIPLE = AGENT BEHAVIOR}

Newell's ontological commitment permitted him to view the knowledge level not only as a convenient method for inquiring about the world but also as a real mechanism by which agents structure data into meaningful patterns used to solve practical problems. Consequently, the effects of knowledge alter the physical world, and the measurable impact of these effects justifies the reality of the level. This allowed Newell to avoid many of the issues that Popper faced because the latter viewed the Rationality

Principle (Popper's version of the knowledge level) as non-material and therefore something unreachable via empiricism and consequently unfalsifiable. In terms of the Intentional Stance, Newell was more clear and stated that, "The intentional stance corresponds to the knowledge level", however, "The intentional stance is unalterably separated from the physical ... there is little doubt that both Dennett and myself are reaching for the characterization of exactly the same class of systems...in particular the role of rationality is central to both" (Newell, 1982, p.34). A more detailed discussion of the similarities between Popper, Dennett, and Newell is provided below

2.0 THE USE OF LEVELS IN COGNITIVE SCIENCE:

In addition to understanding Pragmatism, analyzing Newell's levels also requires comprehension of how levels are employed in cognitive science. This section briefly explores the use of levels as they apply to understanding Newell's theory, for more detailed accounts, see Kersten et al., 2017; Bechtel & Anderson, 2007; Churchland, 1986; Fodor, 1974; Nickles, 1973; and Wimsatt, 1976. The fundamental distinction between how levels are used in cognitive science, psychology, and computer science is the dichotomy between epistemological use and ontological use—though overlap between variations of these two exist. David Marr's Tri-Level Hypothesis and Dawson's slightly modified version of Marr's system are good examples of the epistemological use of levels and they form a foundational basis for how much of the study of human cognition is conducted (Marr, 1977, 1982; Dawson, 2013). As epistemological levels, Marr's and Dawson's systems attempt to provide a framework for inquiring about the world to help us organize our theories about it. These systems also offer a unifying

methodology for the study of cognitive science to better integrate disparate research findings and debates—a goal that Newell himself championed in his famous “You Can’t Play 20 Questions with Nature and Win” paper (Newell, 1973). It is worth pointing out though, that neither Marr’s nor Dawson’s systems employ any version of a ‘rationality principle’, which is so critical to the theories of Newell, Popper, and Dennett.

2.1 Marr’s Levels vs. Newell’s

The beauty of Marr’s Tri-level Hypothesis is that it applies to every cognitive activity (mechanism) that is executed in the brain. Marr’s levels do not map to Newell’s levels—for instance, Marr’s computational level is much broader than Newell’s knowledge level because it covers cognitive activities other than knowledge manipulation. That said, you can apply Marr’s levels to guide inquiry into each of Newell’s levels because in the latter, there are behavior laws, components, and a medium for every level. In Marr’s epistemological system, there is one medium and we use various levels of inquiry to ask different questions about that medium in light of the function/phenomenon under observation. For most researchers who follow Marr, neural realism means if we are talking about brains, then the medium must be neurons. Conversely in Newell’s ontological system, there exists a hierarchy of mediums (one for each level), wherein each one emerges from the level below that medium’s level. Mediums are created when the level below produces a stable and coherent product that can function as a representation of object, entity, or concept. Newell uses computer systems levels to describe his knowledge level specifically because computers are engineered to have very clearly defined levels to divide and conquer the many

complex tasks that take place in computer architectures. These computer systems levels are intentionally designed so that an expert on one level need not understand all the details of another level in order to execute tasks at their own level.

What Newell means by cognitive level—the level out of which the knowledge level arises—is what Marr defined as the ‘procedural level’. While Marr would agree with some of what Newell said, the crucial difference between the two is that Newell thought that there really was a level of cognition devoted to complex manipulation of knowledge in the furtherance of our goals because he saw the impact that knowledge had in reshaping the physical world.

2.2 Epistemological Use of Levels

Marr’s Tri-Level Hypothesis and Dawson’s slightly modified version of Marr’s system are good examples of the epistemological use of levels as frameworks to better order our understanding of cognition and information processing (knowledge manipulation) in human and artificial intelligence systems (Marr & Poggio, 1977; Marr, 1982; Dawson, 2013). Marr’s Tri-level Hypothesis separates all information processing aspects into one of three levels: Computational, Algorithmic, and Implementational ordered from top to bottom (Marr, 1982). Dawson’s system is the same as Marr’s but inserts an Architectural level between the Algorithmic and Implementational levels (Dawson, 2013, p.36). Emphasizing the epistemic nature of these level systems, Dawson writes “Levels do not attempt to explain the nature of information processing devices, but instead provide an epistemology—a way to inquire about the nature of the world”

(Dawson, 2013, p.53). Dawson also argued that the epistemological system of four levels of investigation that he employed was not only useful but *necessary* for explaining any information processor (Dawson, 2013, p. 19).

Although Dawson's epistemic use of levels is not how Newell's levels should be understood, there are some similarities. Like Newell, Dawson explains that each level *emerges* from properties of the level immediately below it. However, each level requires its own unique vocabulary and methodologies. This also means that experts familiar with operating within one level need not necessarily be intimately familiar with the others (Dawson, 2013, p. 19; Newell, 1982, p. 10). Levels are an abstraction of the key details necessary for the functions described in that level and need not contain detailed counterparts to all other levels.

Marr/Dawson's epistemic levels system

Table 1

<u>LEVEL</u>	<u>QUESTION GUIDING INQUIRY AT EACH LEVEL</u>
COMPUTATIONAL:	"What types of information processing (function) problems can be solved by a system?" It transposes descriptions of cognitive phenomena into information-processing problems. This level is focused largely on input-output pairings (in a somewhat Behaviorist fashion). What inputs into the system produce what outputs and how do we use this tool to accomplish goals/tasks?
ALGORITHMIC:	"What steps are required to solve the specific problem?" It articulates the detailed procedure used to solve the problem and organizes it into concise steps that are easy to communicate to other humans or encode into a computer.
ARCHITECTURAL: (Dawson's addition)	"What primitive operations are used as a foundation to the algorithm(s)?" It looks at the components and structure of the algorithm under investigation and does not concern itself with the specific physical substrate that brings the architecture into being.
IMPLEMENTATIONAL:	"How are the algorithm(s) physically instantiated?" It determines which physical mechanisms are responsible for realizing the algorithm(s). Cognitive science pays special attention to this level because it is concerned with multiple realization of the same cognitive processing functions implemented in different substrates—neurons versus silicon chips and relay circuits for example). <i>Multiple realization</i> is the "term used to recognize that different physical mechanisms can bring identical functions to life" (Dawson, 2013, p.36).

It is important to point out that Marr's levels do not map directly to those of Newell. Although Marr's computational level shares much in common with Newell's ideas about the knowledge level, Marr's computational level deals with a much broader range of cognitive activities. Instead of focusing principally on knowledge manipulation as the knowledge level does, Marr's computational level also deals with aspects of memory, perception, 3D internal object rotation, etc. While this paper primarily uses

Newell and Marr's systems levels to help evaluate Newell's Knowledge Level Theory, these two are not the only theorists with systems levels employed in cognitive science—space prohibits a broader discussion of the use of levels in the field of Cognitive Science as a topic unto itself.

2.3 Ontological Use of Levels

The second and less common way that levels are employed in cognitive science and philosophy is the way Newell uses them—ontologically. According to Newell, “A system level is a property of nature, and not just something in the head of the observer” (Newell, 1994, p.118). That is, they define or explain actual states in the world. Newell describes “system levels”—such as those used in computer science and artificial intelligence—as groupings of components that are organized to produce specific behaviors or functions. Newell uses the engineered system levels in computers as evidence that system levels can physically exist (Newell, 1982, p.11-12, see also Causey, 1977).

In Newell's Knowledge Level Theory, each level emerges from the level immediately below it. Newell explained the various elements that comprise a level, “a level consists of a *medium* that is to be processed, *components* that provide primitive processing, *laws of composition* that permit components to be assembled into systems and *laws of behavior* that determine how system behavior depends on the component behavior and the structure of the system” (Newell, 1982, p.10). The claim that brains are built in this way is based on Simon's theory of the evolution of hierarchical systems,

in which he used the parable of the two watchmakers to explain why evolution would tend to produce hierarchical structures (Simon, 1969, pp.188-189). The Knowledge Level Theory further develops the Physical Symbol System Hypothesis that Newell and Simon developed, which is a physically grounded system rooted in Frege, Whitehead, and Russell's research on logic (Newell, 1982, p.11; Newell & Simon, 1976, pp.85-88).

Each level can be defined both autonomously and in the terms of the level below it. Newell explains, "any instantiation of a level can be used to create any instantiation of the next higher level" (Newell, 1982, p.10). The world is effectively organized by the way levels differentiate complex phenomena. Modern science has evolved to the point that we now view the various subfields of the physical sciences through this epistemological perspective of levels, that is, as growing hierarchically out of physics—with chemistry emerging from physics, and biology emerging from chemistry, etc. All systems levels are defined by four key features (Newell, 1982, p. 16):

- **SPECIFICATION OF LEVEL BEHAVIOR:** This determines completely the definite behavior for a system at this level based on initial and boundary conditions (constraints).
- **INTEGRATION:** The behavior of the overall system results from the way in which each level's components process the medium of that level (how a system translates inputs into outputs on its level).
- **CONSTITUENCY/COMPOSITIONALITY:** The infinite possible behaviors of the overall system are a result of its ability to assemble a finite number of primitive components using a set of generative rules. This is based on principles identified by Wilhelm Von Humboldt (1988) and further developed by Noam Chomsky (2017).
- **RELATIVE STABILITY OF MEDIUM:** The medium is realized by state-like properties of matter, which remain passive until acted upon by the level's components (Newell, 1982, p. 16).

Example of the elements comprising two of the systems levels:

Table 2

<u>PROGRAM LEVELS:</u>	<u>SYSTEM:</u>	<u>COMPONENTS:</u>	<u>MEDIUM:</u>	<u>BEHAVIOR LAW:</u>
COGNITIVE LEVEL:	Computers	Memories, Operations	Symbols, expressions	Sequential interpretation
ARCHITECTURAL LEVEL (REGISTER- TRANSFER LEVEL):	Digital systems	Registers, functional units	Bits/bit vectors	Logical operations

Table 2 is merely one example, but you can do the same for the other levels as well—implementational, procedural, etc. Despite the way that levels can be defined in multiple manners, it is vital to recognize how unique each level is. One way to understand this is by looking at the mediums of each level.

Mediums of Computer Systems Levels

Table 3

<u>LEVELS:</u>	<u>DEVICE LEVEL</u>	<u>CIRCUIT LEVEL</u>	<u>LOGIC LEVEL</u>	<u>SYMBOL LEVEL</u>	<u>CONFIGURATION LEVEL</u>
MEDIUM:	Electrons	Current	Bits/bit vectors	Symbolic expressions	Amounts of data (more complex/molecular versions of symbolic expressions)

It is also critical to note that all these computer system levels are approximations or abstractions that are physically realized only to varying degrees and within certain contexts. To keep the consequences of this fact in perspective, it is important to remember that all objects in the physical world are realizations or approximations of some higher general ideal/concept. The earliest version of this idea originated with Plato's idealized forms and continues in our thinking to this day. In this case, our understanding is an approximation of the reality of physical systems, which are in turn only approximations of some higher perfect form. In this structure too, we see the notion of levels emerging from levels.

Newell's philosophical justification for the ontological status of the knowledge level is simply that, as in AI, the knowledge level emerges from the cognitive level below it, "as just another level in the hierarchy ..., there is nothing special about the knowledge level, in any foundational or philosophical sense. (Newell, 1994, p.49).

2.4 Hybridized Use of Levels

Some authors take a middle path on levels that goes beyond epistemology but does not commit fully to an ontological position. Patricia Churchland theorized that the vocabulary and laws of higher level theories provide a more efficient means of referring to far more complex processes at the lower levels (Churchland, 1986)—a benefit Dennett noted about his Intentional Stance (Dennett, 1987, pp.33-35). Likewise, Pylyshyn (1984) distinguishes levels in terms of how they are articulated.

Newell used levels to refer to the actual existence of individual levels within systems (ontologies) and bands to refer to theoretical language appropriate for describing systems levels (epistemologies) (see Table 4). Though not stated explicitly, it appears that Newell may have felt that in cases where highly accurate epistemologies (i.e., those which exhibited a high degree of fit), they could map-directly onto ontologies. Following common practice, I will continue to refer to the neural level, the cognitive level, etc., but it is important to note that these potentially refer to multi-level systems.

Newell's use of levels and bands

Table 4

<u>WORLD</u>	<u>SYSTEM</u>	<u>TIME UNITS</u>
SOCIAL BAND	Project	Months
	Project	Weeks
	Project	Days
RATIONAL BAND	Task	Hours
	Task	10 Minutes
	Task	Minutes
COGNITIVE BAND	Unit Task	10 Seconds
	Operations	1 Second
	Deliberate Act	100 Milliseconds
BIOLOGICAL BAND	Neural Circuit	10 Milliseconds
	Neuron	1 Millisecond
	Organelle	100 μ s

Although Newell used a version of this table to help explain his perspective of levels as they relate to the Knowledge Level Theory, the same system could be

described within the framework of each time unit 1 second, 10 minutes, days, weeks, months, etc.

2.5 Context and Boundary Conditions

Bechtel and Hamilton (2007) can be understood as arguing for the type of hybrid or middle path discussed earlier, wherein they bridge the epistemology-ontology dichotomy. They assert that cognitive constructs, such as symbols or rules, should be considered as ontological entities when their boundary conditions are met, where ‘boundary conditions’ refer to situational and neural constraints that result in the deployment of cognitive structures which function in a consistent and predictable way. Consistent with this view, Newell states, “The medium is realized by state-like properties of matter, which remain passive until acted upon by the level’s components” (Newell, 1982, p.16). This suggests that Newell’s sense of reality is consistent with Bechtel and Hamilton (2007). In addition, the knowledge level can be viewed as providing the boundary conditions of a given problem space or task. In practical/real world problems, these boundary conditions amount to the facts of the agent’s physical environment. In intellectual/scientific problems, they amount to the logical/conceptual constraints imposed by the specific domain of the problem set (which contains a goal) and the relevant knowledge available.

Newell’s distinction between computer system levels and levels in scientific descriptions clarifies what he meant by asserting the reality of the knowledge level. He explained that a computer systems level “does not provide a general closed description

of an entire universe, which is what we generally expect (and get) from a level of scientific description in physics or chemistry" (Newell, 1980, p.12). Instead, computer systems levels such as the knowledge level describe the dynamic relation—series of evolving relational states—between a system/agent and its environment and do not describe the environment itself in its totality. The computer deals with representations/abstracted versions of the environment. Similarly, Dennett noted that organisms' minds continuously mirror their environments and representations of the environment are implicit in the internal organization of the system (Dennett, 1987, p.31). Consequently, computer systems levels are approximations/abstractions and are realized in the physical world (inside the computers when it is turned on) only to various degrees and within specific boundaries of a given environment or medium.

3.0 THE COGNITIVE LEVEL:

Before delving further into the knowledge level, I should first briefly explain the cognitive level because the knowledge level emerges from the cognitive level and, as such, it provides the foundations for accurately understanding the knowledge level. The cognitive level is assumed to be based on neural structures that allow the encoding, storage, transportation, and alteration of neural representations similar to a computer operating system (Newell, 1982, 1990). While this is the predominant view, not everyone agrees with it. There are some theorists—such as Anthony Chemero—who subscribe to a radical version of Embodied cognitive science and believe that no representations exist at all (Chemero, 2009). A cognitive architecture, such as SOAR or ACT-R specifies mechanisms based on these structures, and combinations of these

functions. It is vital to point out that encoding of atomic symbols—the most basic symbols representing external objects inside the mind—happens at the cognitive level and that these atomic symbols serve as the building blocks used in more complex molecular data structures (knowledge) within the knowledge level, where they are manipulated and compiled into ever increasingly complex patterns of meaning.

3.1 Knowledge is Not Found within the Cognitive Level

Knowledge is not to be found within the cognitive level, instead knowledge is an emergent property of the cognitive level and knowledge constitutes the medium for the knowledge level (Newell, 1994). One may ask, what is the value of separating the knowledge level from the cognitive level? The main reason for Newell appears to have been that doing so separates out different problems posed by representation—atomic encoding versus compilation of molecular representations or knowledge for instance—to make them more manageable research programs. The knowledge level is a level in which minimal instrumental rationality is used to move from goals and beliefs (knowledge) to actions—the cognitive level does not contain these types of activities. Within the cognitive level, a representation could correspond to a word or it could also be an internal, non-linguistic symbol or code. Newell defined “symbols” in terms of distal access to memory (Newell, 1994, p.132). This definition means that whenever a mechanism transmits a packet of information to another mechanism, that transmission is symbolic. Effectively, a symbol is a packet of information—which is encoded at the cognitive level—and transported within the system. Thus, cognitive mechanisms

perform actions such as, if representation *A* matches representation *B*, then send representation *A* to location *X*.

3.2 Cognitive Mechanisms Exist when Boundary Conditions are Met

Following from Bechtel and Hamilton's (2007) concept of the real, cognitive mechanisms exist when they are being used (i.e., when their boundary conditions are met within a problem space/task domain) and can be considered "inert" (i.e. they are pure physical symbols without meaning) as Newell and Simon described them, when they are not undergoing processing (Newell & Simon, 1976). Newell and Simon wrote:

"In none of these systems [those of Alan Turing, Emil Post, and Alonzo Church, on which the Physical Symbol System Hypothesis was built] is there, on the surface, a concept of the symbol as something that designates...the data are regarded as just strings of zeroes and ones—indeed, that the data be inert is essential to the reduction of computation to physical process" (Newell & Simon, 1976, p. 89).

Newell took this one step further and hypothesized the existence of an integrated system of cognitive mechanisms, which exists as a cohesive systems level, built on functional neural structures. Newell was clear that this was a hypothesis that needs to be empirically investigated (Newell, 1994). It is not an argument that the cognitive level must exist and must take a particular form. Although Newell did propose a particular theory describing the cognitive architecture (SOAR), his proposed theories about the cognitive level (cognitive architectures) are informed guesses about its actual form.

Also, it is possible that some parts of the brain are not organized so as to produce a cognitive systems level and/or that other types of systems levels exist (Chomsky, 2016). Chomsky's idea of a language module would be an example of a different type of systems level (likewise Chomsky seems to treat symbols as coherent physical entities). Minimally, Newell's hypothesis amounts to an empirical claim that a cognitive systems level exists somewhere in the brain and contributes to determining thought and action.

3.3 Cognitive Modeling Architectures are Task-bound Simplifications

The SOAR architecture and other cognitive architectures influenced by Newell's ideas (such as ACT-R) are often criticized for using overly simplistic symbolic structures and ignoring problems related to representation, meaning, context, and grounding. However, I argue that this criticism is largely due to a philosophical misunderstanding. The knowledge entered into these architectures is derived from a knowledge level analysis of the task. Modelers translate this knowledge directly into symbols for use as representations within the cognitive architecture. This practice can be justified by arguing that it is sufficient to use only the parts of the symbolic representations that are relevant to the task at hand—i.e., those which influence the task's solution. This sidesteps complex issues related to representation, but it means that symbols created in this way should be understood as task-bound simplifications or abstractions. Whether a specific cognitive model is accurate depends on whether it captures all of the relevant details of the task being modeled. Whether the general practice of cognitive modeling is legitimate ultimately depends on how the human cognitive system actually represents

information. However, my point is that the logic for it flows from Newell's concept of the knowledge level as a separate systems level in the brain. With these critical clarifications in place, we can move on to a discussion of the knowledge level in earnest.

4.0 THE KNOWLEDGE LEVEL:

The Knowledge Level Theory expands the number of systems levels commonly used in computer science. Newell defined the *Knowledge Level Hypothesis* in this way, "there exists a distinct computer systems level lying immediately above the symbol [cognitive] level which is characterized by knowledge as its *medium* and the principle of rationality as the *law* of behavior [processing mechanism]" (Newell, 1982, pp. 14-15). For Newell, 'medium' Newell also proposed that the knowledge level exists within the human brain, above the cognitive level. To clarify what Newell meant by 'medium', examples of the pairings of medium and law to systems level are contained in Table 5 (Newell, 1994, p. 47).

Medium and law pairings for levels

Table 5

<u>SYSTEMS LEVEL:</u>	<u>MEDIUM:</u>	<u>LAW:</u>
KNOWLEDGE LEVEL SYSTEMS:	Knowledge	Principle of rationality
PROGRAM LEVEL SYSTEMS: (Symbolic or Cognitive level)	Data Structures, programs	Sequential Interpretation of programs
REGISTER-TRANSFER SYSTEM: (Architectural level)	Bit vectors	Parallel logic
LOGIC CIRCUITS:	Bits	Boolean algebra
ELECTRICAL CIRCUITS: (Implementation level)	Voltage/current	Ohm's law, Kirchoff's law
ELECTRONIC DEVICES: (Implementation level)	Electrons	Electron physics

4.1 Components of the Knowledge Level

Within the boundaries of the knowledge level, we can analyze an agent's knowledge of its environment (beliefs), its goals (desires), and use the principle of rationality to determine the most probable and efficient path that an agent will take through the problem space to reach its goals—this path constitutes the agent's (bounded) optimal behavior. Within the knowledge level, representations are the data structures holding the world knowledge that will be processed into a form that makes a solution available to the agent - the first data structure represents the problem (to include a goal), the second represents knowledge/beliefs (Newell, 1982, p.2). Newell explained that, if a system has a [presumably correct] representation of an object (data

structure) stored in it, then it can be said to possess knowledge of that object (Newell, 1982, p.4). Actions within the knowledge level are processes used to affect an agent's external environment or its internal state in a manner that moves it closer to achieving its desired goal—i.e., progress through the problem space. Dennett and others also note that changes in an agent's environment often also result in changes to its internal state (Dennett, 1987, p.31). These internal state changes can be alterations to beliefs, changes in attitudes/perspectives, etc. that influence the next round of decision-making, hopefully serving to move the agent closer to its intended goal.

The knowledge level is organized around the world of the agent. The cognitive level, which the knowledge sits atop, is organized around the operations available to the agent—particularly as they apply to the encoding of information into symbols. The knowledge level plans out a path to the agent's goals (i.e., is strategic), the cognitive level involves the mechanics of making that path a reality (i.e., the tactical). Newell explained that when you operate in the knowledge level, the medium does not matter—it is always a symbol processing system. Thus, the knowledge level is medium agnostic and can be implemented in a variety of physical substrates (e.g., human brains, computer architectures). Note that this use of medium is like a digital or magnetic media rather than the way Newell employs the term as the key medium of each level (as specified earlier). Thus, the agent (system of the knowledge level) is comprised of:

- **Bodies of knowledge (beliefs).**
- **Goals (desires).**
- **The principle of rationality (constraining logical behavior).**
- **Actions (behavior).**

The knowledge level contains all of the knowledge that an agent possesses, which included the agent's goals (desires) and the information that it has about its environment (beliefs). While knowledge formed the content of this level, the procedure used to manipulate this content was the principle of rationality (the behavior law).

4.2 The Knowledge Level Defines Boundary Conditions for Problem Spaces

The knowledge of the problem provides the *context* or *boundary conditions* of a given problem space or task domain. These boundary conditions amount to the facts of the agent's physical, social, or intellectual environment and include: the constraints of the task, knowledge available, possible solutions, etc. Newell explained:

"An intelligent agent is embedded in a task environment; a task statement enters via a perceptual component and is encoded in an initial representation [at the cognitive level]. Whence scans a cycle of activity in which a recognition occurs (as indicated by the eyes) of a method to use to attempt the problem. The method draws upon a memory of general world knowledge" (Newell, 1982, p.2).

Knowledge Level elements correlated to computer systems level elements

Table 6

<u>LEVEL ELEMENTS:</u>	<u>DESCRIPTION:</u>
SYSTEM:	The system at the knowledge level is the AGENT (as compared with the digital systems at the Register-Transfer level or computers at the cognitive level).
COMPONENTS:	The components at the knowledge level are GOALS, ACTIONS, and BODIES. The agent is composed of a set of actions, a set of goals, and a body.
MEDIUM:	The medium at the knowledge level is, not surprisingly, KNOWLEDGE.
BEHAVIOR LAW:	The PRINCIPLE OF RATIONALITY, which dictates that actions are selected based on their ability to attain the agent's specific goals.

4.3 Knowledge as the Medium

Newell explained that all levels are based on a *medium* that is to be processed (Newell, 1982, p.10). Here, Newell's choice of the word medium may have been less than clear, as it may suggest something that is immediately physical, such as clay for a sculptor or paint for a painter. Neurons have come to be accepted as the foundational medium from which the brain is constructed. When Newell states that knowledge is the medium for the knowledge level it is often interpreted as giving symbols some sort of mysterian existence, independent of the physical nature of neurons. However, this is not what Newell intended. When Newell used the term 'medium', he meant physically instantiated structures that reliably function in particular ways when their boundary conditions are met—such as neural assemblies in brains, or digital medium in computers.

Newell's knowledge level theory rests on the Physical Symbol System Hypothesis, which clearly states that the intelligence systems that Newell discusses obey the laws of physics and are not restricted to human symbol systems of intelligence (Newell & Simon, 1976, p.85). In the case of the medium of knowledge, its ultimate value lies in the fact that the physical symbols being processed are imbued with meaning—i.e., serve as representations for objects in the physical world—i.e., they mirror the physical world internal to the agent (Newell & Simon, 1976, p.89). In this way, meaning is the real medium of the knowledge level, but not in the same way that the meaning is represented at the cognitive level—i.e., that symbol x in the agent's brain stands for object y in the real world. Meaning in the knowledge level is the integration/compilation of symbols into ever-increasingly complex data structures, where two atomic symbols are merged (as in a horse and a bird into Pegasus), or where inferences are drawn from more elementary data representations. Symbolic structures should not be expected to match actual representations fully, as they are abstractions of the relevant details for the specific task at hand.

Dividing the levels in this manner means that some questions about representation and grounding apply to the cognitive level, and not the knowledge level. It permits us to decouple and break down complex problems to be tackled by two different research groups to alleviate roadblocks put up by unnecessary complexity. This has a valuable benefit to AI research as it allows for a divide and conquer or top-down design strategy for dealing with complex representations, starting from a specification at the knowledge level, followed by issues of how to represent the

knowledge, and so on downwards. It also simplifies explaining processes that deal only with manipulation of existing or previously encoded symbols and does not require detailed discussion of how the symbols were encoded and how they are linked to objects in the environment. That said, we are confronted with the problem of how representation actually works when we probe below the knowledge level. For example, the fact that we can answer a question can be predicted by the fact that we know the answer (knowledge level), but the time it takes to retrieve the answer depends on the details of the representation and retrieval mechanisms (cognitive level).

4.4 Impetus Behind the Genesis of the Knowledge Level Theory

Newell's motivation for articulating the Knowledge Level Theory resulted from his concern with the informal and loose manner in which the term "knowledge" was being used in the fields of computer science, cognitive science, and artificial intelligence. Newell attempted to contrast this with the more formal—though far from uniform—way in which the term "representation" was employed (see Brachman & Smith, 1980), despite the fact that knowledge was typically used to refer to the whatever information was encoded and contained in a larger (non-atomic) representation. Newell explains, "If a system has (and can use) a data structure which can be said to represent something (an object, a procedure...whatever), then the system itself can also be said to have knowledge, namely the knowledge embodied in what representation is about" (Newell, 1982, p. 4). Newell saw research on knowledge and representation to be the top priority on the agenda of the field of cognition and especially in artificial intelligence (Newell, 1982, p. 6). Newell explained, "what cognitive science needs ... is a concept of

knowledge that is used to describe and predict the response of functions of [an intelligent] system" (Newell, 1994, p.46). Newell assessed that knowledge was not merely a loose alternate term for representation or logic, he viewed knowledge as a distinct notion for more complex representations or data structures (and rules for generating these more complex data structures) with its own role in the nature of intelligence (Newell, 1982, p. 7).

PART III: Viewing The Knowledge Level Theory through Pragmatism

To fully understand Newell's research, it must be viewed from the perspective of Pragmatism. Viewed through Pragmatism, Newell's argument that 'the knowledge level is real' means 'postulating the knowledge level as real gives explanatory grip to our theories'. It also furthers our understanding of the link between knowledge and action, allowing us to predict future behavior of agents, make sense of their past actions, and plan our own actions accordingly. The goal-directed decision-making that Newell examines and his focus on how knowledge is turned into action make Newell and the Pragmatists natural allies. Consequently, the principles of Pragmatism provide support for Newell's perspective of the reality of the knowledge level.

Some may argue that there is nothing new in applying the philosophic principles of Pragmatism to the study of the mind and the search for artificial intelligence. In fact, Jerry Fodor, in his 2008 book *Language of Thought 2*, notes that Pragmatism is rampant in AI and cognitive science research and he argued that pragmatism as "the leading

source of error in the study of mind today" (Fodor, 2010, pp. 8-12). While I do not completely disagree with Fodor's criticism of pragmatism in cognitive science, I do see a high degree of divergence between the loose terms referred to as pragmatism as it is characterized in the works to which he refers and the philosophical field of Pragmatism's original principles. Perhaps the AI and cognitive science researchers Fodor had in mind have tried to exploit the grounding which they perceive Pragmatism has to offer without fully understanding its actual philosophical implications. This is hard to measure as Fodor did not provide many explicit cases of this pragmatism in AI and cognitive science beyond briefly citing Patricia Churchland (1987). However, we do not have to accept or deny the specific cases that Fodor had in mind because the degradation of the core tenants of Pragmatism began almost immediately, forcing some of its founders—such as Peirce and Schiller—to quickly rebrand their views in order to maintain the clarity and integrity of their terms and principles (Peirce, 1978; Schiller, 2005). With this in mind, the Pragmatism discussed in this paper will be that of early Pragmatists as opposed to these far-ranging subsequent interpretations. Space precludes a full historical discussion of Pragmatist philosophy, instead I will focus on those elements of Pragmatism that are essential to understanding the knowledge level.

Newell's perspective—perhaps without intentionally meaning to—is more philosophically grounded in the original principles of Pragmatism than more overtly 'pragmatic' discussions in AI by virtue of Newell's ontological commitment. Newell employed the principles of Pragmatism in his work, but it is not clear if he would have self-identified as a Pragmatist philosophically. It can at least minimally be attested that

Newell's work is in the spirit of Pragmatism even though he does not regularly cite or advocate Pragmatism specifically. Pragmatism provides a series of commonsense pieces of evidence that support Newell's view that knowledge is real. Viewing theories such as the Knowledge Level theory, the Intentional Stance, and the Rationality Principle through the lens of Pragmatism provides a framework for more equitably measuring and integrating research in cognitive science and AI. It does so because we can more easily determine how each new development moves us closer to or further from our goal, and subsequently, we can identify which elements are relevant to a specific problem rather than trying to understand all complex interactions of a phenomenon even when they do not have a clear bearing on the solution we hope to attain.

5.0 THE PRAGMATIC FRAMEWORK OF THE KNOWLEDGE LEVEL:

5.1 The Origins of Pragmatism

Pragmatism focuses on the practical, observable outcomes of our beliefs, and explains meaning and truth in terms of the effects or impact that believing an idea would have in practice (Barker, 2008). The philosophical field of Pragmatism originated with American philosopher Charles Sanders Peirce. Although Peirce was the founder of Pragmatism, the world is much more familiar with the works of his friends and contemporaries—William James and John Dewey—whose ideas, while still relevant, were largely much easier to digest and consequently more palatable to the broader scientific and philosophical community of the day. Peirce's antagonism toward many of his more Rationalist contemporaries further served to marginalize his thought and sideline him historically. Subsequent thinkers such as Ferdinand Canning Scott Schiller

remained largely true to the core principles Peirce advocated, but further developed these ideas by explicitly tailoring them to a human-centric view of the world. In light of this, Schiller's philosophy in many ways appears to be a pre-cursor some modern Embodied Cognition discussions because it recognized that the unique physiology of human beings shapes the knowledge to which they have access and their search for knowledge is guided by efforts to solve uniquely human problems.

5.2 What is Realism for Pragmatism?

In his criticism that “pragmatism is perhaps the worst idea that philosophy ever had”, Fodor argues that “the mind’s main concern is not acting, but thinking, and that paradigmatic thinking is directed by ascertaining truths...*What minds do is think about things*” (Fodor, 2010, p.8). However, this seems to stop short of explaining the ultimate end of the thinking. What minds do is think about things *in order to determine courses of action (i.e., to solve problems)*. More specifically, thinking serves to find the actions that will attain desired ends. Clearly minds do not think about things as an end in themselves and the things minds principally choose to think about first are the things that will bring their owners closer to their desired goals (ends).

A common criticism of Pragmatism—by Fodor and others—is that the field defines reality by defining it away (Fodor, 2008). This criticism highlights the somewhat counter-intuitive nature of characterizing Newell as a realist while still being Pragmatic. As mentioned earlier, Newell’s realism is a very uncommon version. And some may argue that if you adopt the Pragmatic version of realism, then do you have an actual

realism left or have you defined it out of existence? A counter argument though might ask, if ‘realism’ can be defined away by simply focusing on what is practically relevant in light of what human beings know (the data to which they have access) and what they desire (their goals), then of what viable use is the struggle to define realism? Put another way, if realism can be redefined into a Pragmatic version, then perhaps this brand of realism focuses only on those aspects of information/experience to which human beings have access and are relevant to our struggles to navigate the world in which we find ourselves.

5.3 A Quick Note on Instrumentalism vs. Pragmatism

The term ‘instrumentalism’ is often conflated with Pragmatism, however, this is a case of the confusion Peirce sought to avoid by changing the name of his philosophy. Instrumental was the form of Pragmatism articulated by Dewey, but it has subsequently been adopted to define a wide range of perspectives. The fundamental idea behind instrumentalism is that nothing exists beyond the phenomenon and that all of the scientific mechanisms we observe merely have the function of connecting one observation to another. Instrumental can be thought of as a form of Pragmatism on steroids but our methods are still viewed as explanatory devices. As such, critics of Pragmatism often chose to attack the more radical ideas of Instrumentalism because radical ideas are always easier to target. Because of its more radical nature, there are few true Instrumentalists, but the perspective is still plausible because all we can know is what we have access to from the senses and all the rest is inference.

5.4 *The Real is What Furtherers our Goals*

For the Pragmatists, the real is that which aids in the accomplishment of our goals, which, in turn alters the physical world (Peirce, 1978). The primary value of the knowledge level is its ability to create a space for the formulation of concepts that can be employed and shared to solve problems in an agent's environment that accomplish its goals. The knowledge level is a practical method for organizing and deepening our understanding of our world by providing explanations that permit a greater degree of accuracy in predicting the behaviors of agents and phenomena in our world, which permit goal attainment—the most vital of which is survival.

The knowledge level itself does not seem intuitively physical to most people, however, it is always physically instantiated and its impact upon the physical world is felt and consequently can be measured indirectly via its effects. Similar to the force of gravity—which cannot be seen, touched, or heard, yet exerts a measurable force on physical objects—the force of knowledge directly impacts the physical world. This view is consistent with Peirce's Pragmatic maxim—"Our idea of anything *is* our idea of its sensible effects...consider what effects, that might conceivably have practical bearings, we conceive the object of our conception to have...then, our conception of these effects is the whole of our conception of the object" (Peirce, 1978, p.31). Within this perspective, the practical effects of a belief or knowledge concept are indelibly linked with that concept's reality. Consequently, the verifiable physical effects of knowledge concepts —buildings arising from blueprints, lives saved through medical breakthroughs, etc.—justify the knowledge level's reality. Put another way, if the

physical effects (impact) of knowledge concepts are real, then so too are the sources of these effects. Thus, the reality of the knowledge level is evident by the solutions its general concepts provide to human problem-solving and this can be measured through the resultant actions that reshape the physical world in the furtherance of human goals. Peirce asked, "What do we mean by real...the real, then, is that which, sooner or later, information [knowledge] and reasoning would finally result in...since no cognition of ours is absolutely determinate, generals [general concepts] must have a real existence" (Peirce, 1978, p.247). While this does not look like most commonly accepted forms of realism, it nonetheless exhibits a certain logic—as goes the reality of the effects, so goes the reality of the cause.

Peirce notes that our modern conception of force in physics arose in the 17th century from the crude idea of a cause and has been constantly developed to the point that it has "given birth to modern science [of which physics is the foundation] and changed the face of the globe...has played a principal role in directing the course of modern thought, and in furthering modern social development" (Peirce, 2016, p. 33). Peirce asked, "what is the immediate use of thinking about force; and the answer is, that we thus account for changes of motion", which has provided knowledge that humankind has used in to make numerous predictions and which has greatly enhanced our understanding and ability to reshape the world (Peirce, 2016, p. 33). Based on the broader context of his writings on this issue, Peirce does not appear to oppose making inferences from the things we can see (observable impacts) to the things that yet remain beyond our reach—see his discussions on abduction and hypothesis (2016) and

his characterization of reasoning as moving from what we know to what we do not know (1978, p.7). While this claim that the content of our conceptions is largely bound up with their resultant effects was an important claim for Peirce, it was also controversial for some.

In this way SOAR can make decisions in circumstances it has not encountered before (draw inferences) and learn from the feedback. This is consistent with Peirce's characterization of reasoning, "the object of reasoning is to find out, from the consideration of what we already know, something else which we do not know" (Peirce, 1978, p.7).

Peirce's point was that the utility of the general concept of force (our knowledge of it) is that it permits us to better explain and predict the behavior of objects in the physical world. The parallel with Dennett's differentiation between the Physical Stance and the Intentional Stance is clear as is the similarity with Popper's Rationality Principle—i.e., that we use these concepts and perspectives to help us better predict the behaviors of physical objects (via the physical stance) and intelligent agents (via theories like the Knowledge Level theory, the Intentional Stance, and the Rationality Principle) to achieve our desired ends. In many ways, the knowledge level has been treated with a similar sort of mysterianism as that which Peirce described about force. While we all can see and agree upon the effects of knowledge—the myriad ways in which it reshapes the physical reality of the world—concerns remain regarding the existence of knowledge concepts and the knowledge level itself.

5.5 Knowledge Transforms the Physical World

Schiller's brand of Pragmatism took a strong stance on the ontological reality of ideas/concepts (knowledge). Schiller argued even more fervently that the power of cognition actually transforms the physical reality of the world and consequently our desires (goals) and ideas (beliefs) are real by virtue of their impact on the physical world (Schiller, 2005). Philosophy historian Julien Marias wrote that Schiller viewed human desires and ideas (knowledge) as real forces in the configuration of the world (Marias, 1967). Through their exploitation of knowledge, human-beings have a significant impact on their environments by continually reshaping the world to affect the primary goal of ensuring their survival. Most other animals focus their energies on adapting to an existing environment rather than transforming it. In fact, no other animal expends as much time and energy transforming its environment into what it wants it to be as do humans. The ability and urge to reshape our world is a significant part of what it means to be human. It is directly made possible by our unique biological makeup—opposable thumbs, advanced neocortex, and complex language capabilities. Particularly in the field of computing and AI, each new goal attained enables greater and greater manipulation of knowledge which, in turn, fuels the next round of computation and innovation to solve the next goal.

5.6 Knowledge is Acquired by Doing

Peirce also shared Newell's experiential perspective that we acquire knowledge as participants—by doing or through experience, as the Empiricists endorsed. Peirce argued that knowledge is an activity, not a spectator sport. This view has much in

common with the concept of forward engineering that fueled Newell's artificial intelligence research, where "doing" often comes before and enables "understanding". Newell also offered his own work—on the Physical Symbol System Hypothesis, the SOAR cognitive architecture and the Knowledge Level Theory—as further evidence of how conceptual advancement grows out of the daily practice (Newell, 1982, p. 8; Newell & Simon, 1976). Peirce emphasized the strong causal link between knowledge and experience by noting its evolutionary value to our species. HE said, it is "Because we are part of the living world, and it is primarily through the pursuit of survival [humankind's fundamental goal] that we acquire knowledge...the most valuable thing about knowledge is its explanatory power" (Peirce, 1978). Thus, we see self-preservation as the original goal setting the chain of our reasoning and behavior in motion.

5.7 Everything Experienced Must Somewhere be Real

James focused on exploring how an agent's unique collection of beliefs (knowledge) was derived from its experience within its environment. James outlined his famous *metaphysical first principle* as "Every kind of thing experienced must somewhere be real" (James, 2000, p.27). An agent experiences an idea similar to the manner in which it experiences a sunrise or the feel of a tree's bark. As an idea appears in the agent's mind—through its own discovery or via its being imparted by another—the agent has an experience of it. The agent turns it over and over in an attempt to understand it, and the agent changes aspects of the idea to enhance it and to make it applicable to a broader range of problems and to make it better fit into the agent's broader view of the world. In this way our knowledge (beliefs) continually evolves and

this process is accelerated when it is written down and shared. James believed that knowledge (belief) was *true* only to the extent that it aided in an agent's establishing satisfactory relationships with all other aspects of the agent's experience—that is, the truth value of a particular piece of knowledge/belief was dependent on its consistency with the larger body of our world knowledge (James, 2000; Marias, 1967, p.397). This urge for consistency is what helps maintain our world view, the attitude with which we approach new knowledge and by which our resistance to new knowledge is determined. This world view is shaped by experience of both our physical and social environments.

5.8 Application of Pragmatism to the Theory of Knowledge

Schiller is perhaps an even greater champion of the ontological nature of general concepts (knowledge) than other Pragmatists as he broadened the application of Pragmatism beyond purely scientific descriptions of our world to include pursuits within the Humanities. Schiller noted that Pragmatism was “a special application of Humanism to the theory of knowledge...but Humanism will seem more universal...It will seem to be possessed of a method which is applicable universally, to ethics, to aesthetics, to metaphysics, to theology, to every concern of man, as well as to the theory of knowledge” (Schiller, 2005, p.16). Our heavy reliance on the use of shared knowledge to navigate our world is one of the more unique facets of human nature. This is because it underpins our scientific endeavors, to which our species has devoted a significant portion of its time and resources. Schiller’s “human experience” is precisely the type of data to which Newell refers as knowledge in the systems he, Dennett, and Popper all describe. Despite being written in 1907, the applications of Schiller’s method to the

study of human cognition and artificial intelligence—in particular cognitive modeling—are abundant. For example, Schiller's work helps bridge from the purely scientific descriptions of phenomena (which make up a significant part of collective knowledge) to the assignment of values for certain patterns of behavior that are the domain of the Humanities. This issue is currently of immense interest as evidenced by the attention that Jordan Peterson's work is receiving (Peterson, 1999). In this way, Schiller's work provides insight for bringing the discussion of values into the rudimentary goal-directed decision making of cognitive modeling, which can enhance understanding of how individual goals are selected, adjusted, deconflicted, and prioritized as new information is made available via an agent's interaction with its environment.

5.9 The Knowledge Level in the Scientific Method

Recognizing the limitations of humankind's access to an objective truth, Pragmatists viewed 'truth' as a process of incrementally enhancing the utility of our explanations to come closer and closer to optimal solutions—vice attaining awareness of some absolute truth. Schiller noted, "All testing of 'truth' is, therefore, fundamentally alike...it always implies an experiment...and it always ends in a valuation" (Schiller, 2005, p.7). These valuations are determined by how the specific 'truth' (knowledge) enhances our ability to solve problems. This is consistent with Popper's theory of scientific progress—which is shared by many in modern science. Popper described scientific progress by noting that, "science always begins and ends with problems...the progress of science lies, essentially, in the evolution of its problems" (Popper, 1994, p.155). Popper viewed the evolution of problems as the sharing, empirical testing,

analytical debate, and the final collective assignment among the scientific community of a valuation of one theory over another based on its practical ability to explain more than the previous theory. Where Schiller and Popper diverged was the in fact that Popper differentiated between scientific problems and practical problems. Popper said, "scientific problems [searches for truth] are preceded, of course, by pre-scientific problems, and especially practical problems" (Popper, 1994, p.156). Schiller viewed all such searches for truth (especially those in science) as having a practical nature because he argued that "all mental life is purposeful" (Schiller, 2005, p.5). The knowledge level is the space in which this collective evolution of science takes place.

5.10 Context Determines Truth

Similar to Bechtel (1985) and Newell (1982, 1990), Pragmatism too defined all truth/reality as dependent upon the context or boundary conditions of a situation. In his analysis of Dennett's intentional stance, Bechtel argued that observers of an agent should adopt a realist perspective towards intentional attributions but should do so within the confines of the conditions of the agent's environment by treating such attributions as relational properties (Bechtel, 1985, p.473). Peirce said, "there are phenomena within our own minds, dependent upon our thought [i.e. dependent upon their relation to our thought], which are at the same time real [objective] in the sense that we really think them...but though their characters depend on how we think them [is relationally-dependent], they do not depend on what we think those characters to be...thus a dream has a real existence" (Peirce, 1978, p.36). That is, the phenomena in our minds of which Peirce speaks are real, but real in relation to our thought.

As we will see in more detail later, Dennett too argued that patterns of belief had objective existence (Dennett, 1987, p. 25) and “these are the patterns that we characterize in terms of the beliefs, desires, and intentions of rational agents” (Dennett, 1978, pp. 27). Dennett explained, “my thesis [the Intentional Stance] will be that while belief is a perfectly objective phenomenon...it can be discerned only from the point of view of one who adopts a certain predictive strategy, and its existence can be confirmed only by assessment of the success of that strategy” (Dennett, 1987, p.15). More pointedly, Dennett argues that “the intentional stance description yields an objective, real pattern in the world” and he feels that this is “confirmed beyond skepticism” (Dennett, 1987, p. 34). What Dennett means by this is that the ‘objective real pattern’ is real by virtue of the fact that the intentional stance reveals it as being approximately similar across individuals interpreting the data and recognizing the pattern.

Schiller argued that ‘abstract’ truths were not real truths because truth was essentially dependent upon context of the problem or phenomenon on which the truth shed light (Schiller, 2005, p.8). In the case of the knowledge level, the context amounts to the boundary conditions of the agent’s environment or those of the problem set at hand. Satisfying Schiller’s own requirement for context/boundary conditions therefore, would amount to ensuring that we maintain the relevant details in our abstraction of a given problem being solved. The essence of the Pragmatic method, according to Schiller, is that “the meaning of a rule lies in its application...that is, it rules and is true within a definite sphere of application [within certain boundary conditions], which has been marked out by experiment” (Schiller, 2005, p.9). Schiller explained for a statement

to be true, it had to be tested by being applied, only after which could we determine its real meaning and what conditions must be met for it to be real/true (Schiller, 2005, p.8).

6.0 PRAGMATISM'S GROUNDING IN EMPIRICISM:

Because, as we have seen, the Pragmatic test of whether something is real is a measure of its being useful and that use must be measured through testing its ability to aid in solving a problem, we see Pragmatism is strongly grounded in empiricism.

6.1 Similarities between Newell's Knowledge Level and Empiricism

Although Newell's assertion of the reality of the knowledge level—which includes the reality of general concepts—may seem at odds with more traditional views of Empiricism, he has key ideas that are similar to Empiricist thought. The following are a few of those similarities.

Newell follows the generally accepted Empiricist view that sensory interaction with the world (experience) results in knowledge (Newell 1982, 1990). Both Newell and the Empiricists viewed experience as the primary source of knowledge for an agent—for Newell this meant the source of the internal representations stored inside the agent over time that comprised its world knowledge. The agent's body of knowledge is comprised of facts about its environment which are acquired directly through interaction with its environment as revealed through the five senses. Truth, Locke argued, consisted of propositions about the state of the world and knowledge was used to determine the truth or falsity of specific propositions (Locke, 1999). The specific

truth value of a proposition or piece of knowledge lay in its consistency with empirical data.

Although most scientists rely upon empirical observation to test the utility of a given theory, Pragmatists, such as Peirce and Schiller, asserted that the truth of any statement could only be defined within a given context or, as Newell would have termed it, within a “Problem Space” (Pierce, 2016; Schiller, 2005). James’ own definition of truth involved a heavy reliance on how new propositions were consistent with one’s larger body of existing knowledge gained through prior experience (James, 2000). The agent employs this knowledge to analyze its options for action and uses the rationality principle to select the most efficient means of accomplishing its goals in that specific environment. When actions are taken but goals are not attained, this too produces new knowledge, which can still lead to internal changes in agent (adjustment of its internal representations that mirror the environment). The agent again uses its senses to ascertain whether its actions changed the state of its environment, which also produces additional knowledge, and then the cycle begins anew—although now with additional knowledge that the agent did not possess during the previous round (Newell, 1982).

Although it is commonly accepted that the mind plays a crucial role in organizing the sensory data collected from the agent’s into various structures which are then used to determine appropriate actions to take in order to achieve desired states (goals), Newell developed this idea further by explaining that the knowledge level is the

workspace within which knowledge gets compiled, combined, evaluated, and re-used to solve ever-increasingly complex problems—i.e., the organization of our sensory data (Newell, 1982).

7.0 NEWELL'S MODEL INSTANTIATES PRAGMATISM:

In his biographical memoir of Newell, Simon noted that Newell devoted his life to answering the question ‘what is the mind?’ . Simon explained that Newell “*gave body to this dream*” and that Newell “had a vision of what human thinking is. He spent his life enlarging that vision, shaping it, *materializing* it in a sequence of computer programs that exhibited the very intelligence they explained” (Simon, 1997). In this singular and heart-felt explanation of Newell’s life, we see the very epitome of what Newell described in his 1982 paper “The Knowledge Level”—namely that our immaterial ideas breathe life into the world and shape that world in physical ways that can be felt and measured. In Newell’s case, this was to model the mechanisms of the human mind—to take his theories of cognition and create a physically-instantiated simulation that showed how they worked.

7.1 Newell’s Model of the Brain as a Pragmatic Machine

Newell’s computational work was not only clearly Pragmatic in nature, his model of the way the brain works argues that the brain is a Pragmatic machine—that is, it selects from among various options of actions those which move the agent closer to attainment of its goals, within the context of the knowledge it has about the problem set (aspects of its environment). Newell wrote little on his philosophical beliefs, but he

and Simon created a normative model of intelligence, the Problem Space (Newell, 1972), and Newell co-designed one of the first cognitive architectures, SOAR (Laird, Newell, Rosenbloom, 1987).

I argue that Newell's philosophical beliefs are implicitly embodied in these two projects. If one looks at Newell's research on problem space and the SOAR architecture, a clear central theme of Pragmatism is apparent—namely that all action is defined in service of a goal (Newell, 1982; Newell, 1990). This view is consistent with Pragmatism's definition of truth/reality discussed earlier. Newell and Herb Simon explained, "All information processed by computers is in the service of ends [goals] and we measure the intelligence of a system by its ability to achieve stated ends in the face of variations, difficulties, and complexities posed by the dynamic task environment" (Newell & Simon, 1976). Newell saw the goal-oriented nature of production systems as the means for creating this type of intelligence in the SOAR architecture.

7.2 Computational Expressions of Pragmatism's Principles

Problem spaces interpret problems as having a starting node (or state), a goal node, and interviewing nodes that bring the agent from the starting node to the goal node. The agent traverses the nodes by making decisions that determine actions taken based on the information available at the current node which results in taking it to a subsequent node. However, at each node there are typically several different possible actions leading to different nodes. Problem-solving is therefore understood as knowing what to do next based on the information at hand (i.e., the current node). The SOAR

architecture navigates problem spaces by using production ('if this-then do this') rules. At each node, only the productions that match the information associated with the node can "fire" and execute the actions required to move to another node. Each production rule is associated with a specific utility value. Experts have finely-tuned utility values such that in the event that more than one production matches the firing conditions, the highest utility production is chosen, and the agent will move to the best, next node leading to the goal. If no production matches, then the system has reached an impasse. To deal with this, SOAR will compare the information associated with the node to other nodes it has encountered in the past, select the most similar, and see if there is a production for that. In this way SOAR can make decisions in circumstances it has not encountered before (draw inferences) and learn from the feedback. This is consistent with Peirce's characterization of reasoning, "the object of reasoning is to find out, from the consideration of what we already know, something else which we do not know" (Peirce, 1978, p.7).

From the perspective of the above Pragmatic principles, it becomes clear how Newell's research projects on Problem Spaces and SOAR are computational expressions of the fundamental beliefs of Pragmatic Philosophy—that knowledge is true only within the context of a specific goal and that solving of problems guides an agent's actions. Applying Marr/Dawson's levels to Newell's research, we see that navigating to a goal is the computational problem, Problem Space is the algorithm, SOAR is the architecture, and neural systems for storing and processing symbolic information is the substrate or physical instantiation.

PART IV: The Knowledge Level as a More Committed Perspective:

Newell was not the first to articulate a theory to describe the cognitive mechanism underpinning our ability to attribute intentions to and predict the behavior of agents. However, Newell's insistence that this mechanism had its own real existence instead of simply being an epistemology avoided limitations suffered by other theorists. The case for this can be made by comparing Newell's Knowledge Level theory to the more well-known theories of Popper's Rationality Principle and Dennett's Intentional Stance. Newell, Dennett, and Popper all agree that these theories can be seen as an effective epistemology for explaining the ability of humans to predict the behavior of intelligent systems (agents). What links Popper, Dennett, and Pragmatism is the idea that absolute truth is beyond human understanding—what we can do is come up with solutions and theories help solve deeper and deeper problems (Popper, 1994, p.151). Newell's theory, however, goes further by claiming that the mechanism these theories explain also has ontological existence.

Newell believed that the knowledge level mapped directly onto the actual ontological structure that predicts and guides agent behavior. This assertion means that the Knowledge Level theory is an epistemology for understanding this mechanism, but the mechanism itself—which Newell characterized as the knowledge level—is part of the ontology of the world. This assertion allows Newell to avoid seeming points of confusion that Popper and Dennett ran into with their own theories regarding

falsifiability and the relativity of intentional attributions. This section explores how the theories of Popper, Dennett and Newell were all variations of descriptions of a complex cognitive mechanism that permits intelligent agents to predict the behaviors of other agents and plan their own actions accordingly to attain their goals. It also highlights how Popper and Dennett unnecessarily limit the value of their theories when they deny the reality of this mechanism.

8.0 WHY DOES THE ONTOLOGICAL NATURE OF THE KNOWLEDGE LEVEL MATTER?

Although all three theories discuss the mechanism by which intelligent agents predict one another's behavior—and plan their own accordingly—Newell is the only one who attempted to deal with this mechanism in a precise computational manner. The ideas contained within Newell's Knowledge Level theory formed the basis for his cognitive architecture—SOAR, which was a tool he used to test experimental theories through simulation. Neither Popper nor Dennett developed a system for simulating their theories through models. Newell's having done so helped test the accuracy of his theoretical conjectures through simulations and better integrate them into existing research programs in computer science and artificial intelligence.

Newell viewed the reality of the knowledge level as being critical to understanding knowledge representation in agents. Knowledge representation in Newell's system consisted of: 1.) "the data structures which hold the problem (goal) and will be processed into a form that makes the solution available" and 2.) "the data structures

that hold the world knowledge”—that is, the agent’s knowledge of its world (Newell, 1982, p.2). He also saw it as vital to moving from a convenient way to ask questions about this predictive capability in human cognition to a way to model theories more precisely through computer simulations to test (falsify) and refine them (Newell, 1982). The field of cognitive modeling emerged from Newell’s effort to integrate and add precision to the myriad theories of cognitive psychology and the search for artificial intelligence (Newell, 1990; Anderson 1983).

9.0 POPPER’S RATIONALITY PRINCIPLE:

Describing the mechanism that Newell characterized as the knowledge level, Popper argued that it was a mistake to assume that the specific individual human psyche was the animus behind a social model of behavior. Instead, he argued that knowledge and aims (goals) were the abstract and typical situational elements that could be used to replace the role previously played by conscious and unconscious psychological experiences to provide the animation of social models—or the series of actions/behaviors of an intelligence agent with its environment and interactions with other agents (Popper, 1994, p. 168). In addition to knowledge and goals, Popper explained that, “there is only one animating law involved—the principle of acting appropriately to the situation, which is clearly an *almost empty* principle...it is known in the literature under the name ‘rationality principle’ ”(Popper, 1994, p.169). He argued that situational analysis (of past actions) and prediction (of future actions) of intelligent behavior actually required very little. Popper noted, “Once we have constructed our model of the situation, we assume no more than that the actors act within the terms of

the model, or that they ‘work out’ what was implicit in the situation...this, incidentally, is what the term ‘situational logic’ is meant to allude to” (Popper, 1994, p.169).

Thus, we see that Popper outlined a formula similar to Newell’s—namely that an agent’s unique combination of its knowledge of its environment and goals are the variables that determine the agent’s rational behavior within the context of a given situation. The constant in this equation was the rationality principle. Popper explained that “we must remember, of course, that the situation, as I use the term, already contains all of the relevant aims and all the available relevant knowledge, especially of the various possible means of realizing these aims” (Popper, 1994, p.169). Popper is referring to what Newell called the “problem space” (Newell, 1972).

9.1 Popper’s Criteria for Determining the Scientific Nature of Theories

Popper’s explained that progress in scientific discovery “depends on instruction and selection; on a conservative or traditional historical element and on a revolutionary use of trial and the elimination of error by criticism, which includes several empirical examinations or tests—that is, attempts to probe into the possible weaknesses of the theories, attempts to refute them” (Popper, 1994, p.7). He deemed this process “falsifiability” and argued that any theory which could be considered scientific must meet these criteria.

9.2 Contradictions with Popper's Rationality Principle & Falsification

A key distinction between Newell's theory and that of Popper turns on the latter's careful emphasis that the rationality principle was "almost empty" or a theory without empirical content (Popper, 1994, p.169). Being without empirical content also put Popper's rationality principle theory outside the realm of falsifiability, which according to his own demarcation relegated it to pseudo-science. Popper regarded the principle of rationality as an empirical conjecture rather than a methodological principle. He viewed it as highly useful and an "integral part of every, or nearly every, testable social theory", but not something that possessed existence in itself—i.e., not real (Popper, 1994, p. 177). Although Popper believed the rationality principle was a valuable way to organize our thoughts on human cognition (i.e., an epistemology) to make predictions about an agent's behavior, he did not view it as real because it was not tangible/physical and was therefore not something that could be subjected to empirical observation and measurement. Thus, we can see that both Popper and Newell saw practical value in this principle, but Newell's deeper commitment to Pragmatism allowed him to treat his knowledge level theory as an ontological reality whereas Popper's decision to continue to view it epistemologically limited his being able to consider it real science (Popper, 1994, pp.177-178).

According to Popper's own theory of scientific discovery, if something could not be measured and tested via empirical observation, then it was not falsifiable and consequently was not scientific. Popper argued, "all the statements of empirical science (or all 'meaningful' statements) must be capable of being finally decided, with respect to

their truth and falsity; we shall say that they must be ‘conclusively decidable’...this means that their form must be such that to verify them and to falsify them must both be logically possible” (Popper, 1994). To summarize Popper’s position, if the Rationality Principle (knowledge level) refers to a system that is not actual/real, then it cannot be subjected to empirical study in order to test to see if it is false and because falsifiability of a theory is required for that theory to rise above the level of pseudoscience to become actual science. Therefore, we see that Popper is stuck.

Although Popper emphasized the need for scientific theories to be falsifiable, he was willing to entertain the idea that some theories may not be falsifiable by the methodologies currently available but could potentially be falsifiable at some point in the future. This flexible forward-looking perspective is shared by Dennett as well as he assessed that once our understanding of the physiological structures of the brain improved, we would be able to map these structures to the intentional stance (Dennett, 1987). This is a key area for further exploration of theories such as the knowledge level without dismissing them out of hand as pure pseudo-science because they have proven utility. As both Newell (1982) and Dennett (1987) pointed out, we employ the cognitive mechanism behind the knowledge level/intentional stance on a daily basis with a high degree of accuracy in its predictions.

9.3 Popper's Process of Scientific Discovery

Running counter to traditional Empiricism, the first essay, “The Rationality of Scientific Revolutions”, in Popper’s 1994 book *the Myth of the Framework* argues that

there is no such thing as neutral empirical observation—all observation is biased because it is preceded and shaped by expectation of some sort—that is by our initial theories. Popper argued that the objectivity of science resides in how initial conjectures (predictive theories) are tested and found to map to observable realities. He felt that scientific theories were “better and better solutions of deeper and deeper problems”, where ‘better and better’ meant “nearer and nearer to the truth”—thus Popper believed there as a clear way to measure scientific progress (the collection of validated knowledge). (Popper, 1994, pp. 154-161). The measure for scientific objectivity and progress that Popper outlined possesses a very practical nature and its being tied directly to problem solving makes it a good macro example of the type of activity Newell described as taking place within the knowledge level. In many ways, the scientific community is a distributed version of the problem-solving processes that Newell described as taking place inside the individual.

Peirce too believed that science started with a creative theory, “When a man desires ardently to know the truth, his first effort will be to imagine what truth can be...there is, after all, nothing but imagination that can ever supply him an inkling of the truth. He can stare stupidly at phenomena; but in the absence of imagination they will not connect themselves together in any rational way” (Peirce, 2016, p.43). Arguing for the somewhat mystical nature of the creative aspect of science, Peirce said, “Find me a people whose early medicine is not mixed up with magic and incantations, and I will find you a people devoid of all scientific inquiry” (Peirce, 2016, p.43).

According to Popper, the objectivity of science lies not in the individual scientist or his/her theory, but in the extended process of critical discussion of a theory(s) among the community of scientists over time and the resultant modifications of theories when conflict with empirical observations is identified (Popper, 1994, p.159). For a theory to exhibit objectivity, it must have been subjected to broad criticism. This criticism is the primary means by which science progresses. Popper was careful to point out that the only function which his theory of scientific method assigned to empiricism (observations, experimentation, and measurements) was the “modest although important one of assisting criticism—that is, in assisting in the discovery of our mistakes” in order to bring them more in line with observations of the real world. Consequently, the empirical content of any theory is a factor of whether it is refutable and can be subjected to criticism whose arguments are supported by the weight of empirical observation (Popper, 1994, chapter 8, section 1, p. 162). That is what Popper meant by the “falsifiability” of a theory.

9.4 Popper's Scientific Process Compared with the Knowledge Level

If empirical observation played a limited role in Popper's view of how science progresses, then the main arena of activity must be in the larger conceptual dialogue among scientists regarding the relevant strengths and weaknesses of specific theories. The theories that Popper talked about are themselves, from Newell's perspective, data structures that exist—and are transmitted from one person to another—within the knowledge level. Thus, even in Popper's system the primary factor in scientific discovery takes place within what Newell defined as the knowledge level because it

deals with the sharing of concepts that aid in providing generalized descriptions of physical phenomena. Humans then, in turn, use these concepts to plan ways to reshape their physical environment.

Steps of Popper's Scientific Process: Popper (and many others) defined the process of scientific discovery as comprising the following steps. The following bracketed items show how Newell's theory accounts for the various elements of this commonly accepted description of the scientific process.

1. Start with a clearly defined problem set or phenomenon [knowledge level beliefs that define the problem space/task domain to be solved in the environment].
2. Begin to make conjectures and theories as to why this phenomenon came about, and/or how it works [assemble bits of knowledge/data into constructions that help solve sub-goals and progress through the problem space].
3. Test the individual conjectures and larger theories against empirical observations/data [compare our theoretical prediction to empirical data to test their fit]. This not only helps to test theories for accuracy and completeness, but it also helps clarify our terms and deepens our understanding of other factors not originally considered. It can also lead to discovery of new phenomena/problem sets, which in turn sets the systems into the next cycle of scientific development. [For Newell, this would represent the process of the agent's trial and error being catalogued as new knowledge and used to enhance the next round of decision-making to move the agent through the problem space towards its goal].

The primacy that Popper grants to initial theories—constructs within Newell's knowledge level —over empirical findings can be seen in the twist he offered on the empiricist tradition that began with Francis Bacon by emphasizing that observation was not the initial basis for forming our theories. Popper noted that empirical observation of nature does not instruct us in how to build theories, but instead we begin by forming

theories/conjectures/inferences drawn from our existing knowledge. We then use empirical observation to test—corroborate or disprove—those theories based on whether or not they fit the facts of the world. If our theories do not fit well, they are adjusted to better match empirical findings. These adjusted theories go on to guide the next round of empirical testing in an ever-increasingly complex cycle until the theory is rejected or replaced. Popper goes so far as to say that we cannot make even simple observations without some sort of theory through which to interpret our findings (Popper, 1994, chapter 8). The initial theory is necessary to frame our investigations/observations. The entire process that Popper describes as scientific discovery is effectively a process of building predictive strategies to explain the behavior of external objects based on what we know of their nature. This process takes place within Newell's knowledge level.

9.5 The Rationality Principle's Pragmatic Goal-Directed Approach

Popper's essay, "Models, Instruments, and Truth: the status of the rationality principle in the social sciences" highlights the pragmatic nature of his search for truth as solution to meaningful problems. Popper wrote that "science always begins and ends with problems" and that "the progress of science lies, essentially, in the evolution of its problems" (Popper, 1994, p. 155). Popper argued that practical problems precede scientific problems, and these problems arise when our expectations are disappointed—or in Newell's vocabulary when our planned course of action fails to achieve our goals. When Popper points out that "we never start afresh...the growth of knowledge always consists in correcting earlier knowledge", what he is effectively saying is that scientific

progress is the ever-increasing sophistication of our collective knowledge base about our environment and the problems it poses to us as agents (Popper, 1994, p. 156). Consequently, Popper viewed specific items of knowledge—a scientific theories for example—as a tentative solution to some specific practical problem faced by human beings (Popper, 1994, p. 156). The sophistication level is a measure of how well the new knowledge answers our scientific questions and thus solves our practical problems. Although we may ask purely theoretical questions along the way, the initial motivation for our exploration of theoretical concerns is almost always kicked off and guided by a practical problem. Popper thus viewed understanding a theory as viewing it “as an attempt to solve a certain problem” and that “without understanding the problem situation which gave rise to the theory, the theory is pointless” (Popper, 1994, 157). Thus, like Newell, we see that knowledge is indelibly linked with our effort to find solutions to achieve the problems that beset us as we move through our environment. Or, put into Peirce’s terminology, it is the practical consequences of the theory or general concept that imbue it with purposes and meaning, and thus reality.

9.6 Distinction between Popper’s Theory and Newell’s

Despite the highly practical nature of this perspective, Popper considered himself opposed to the type of instrumentalist philosophy of science proposed by Pragmatists such as Peirce and John Dewey, though it is clear that at times, he shared many aspects of Pragmatism in his views (Popper, 1994, p.154). Although Popper and Newell’s respective theories share many similarities, one point distinguishes their perspectives. This key difference was motivated by Popper’s strong adherence to a

search for objective truth. Popper believed that an individual theory or conjecture (i.e., a piece of knowledge) was never proved or settled, or established, but rather that sound theories were those which had been subjected to efforts to falsify/refute them that failed. He felt that our hypothetical or conjectural theories were continually revised, modified, and improved, through our attempts to falsify them thereby bringing us closer to some greater truth, whose value was measured by its ability to explain more of a phenomenon than the previous theory did or to allow for greater consistency with other peripheral theories (Popper, 1994).

10.0 DENNETT'S INTENTIONAL STANCE:

The critical role granted to intentionality that is common in the field of cognitive philosophy largely originated from the work of Franz Brentano (1874) who first explored intentional systems. Intentional systems are those systems capable of possessing thoughts, attitudes, expectations, desires, etc. *about* the world—in short, what Newell termed ‘agents’. This definition was used to clearly delineate mental phenomena from their physical counterparts to better frame discussions on the various forces at work in the world. We see in the importance that intentionality gives to mentalistic terms the counter balance to Empiricism’s hyper focus on the purely physical aspects of observable phenomena. Leveraging the concept of intentionality, philosopher Daniel Dennett developed a theory that describes the same cognitive mechanism that humans use for predicting the behavior of intelligent systems, which is in many ways similar to Newell’s knowledge level. Dennett called this view the Intentional Systems Theory or the “intentional stance” (Dennett, 1971, 1987, 1991, 2009).

10.1 The Role of Intentionality

Building on Brentano, and subsequent philosophers, Dennett defines his Intentional Systems Theory (intentional stance) as an analysis of the meanings of intentional ascriptions—that is, everyday “mentalistic” terms such as ‘believes’, ‘intends,’ ‘expects’, ‘desires’, ‘decides’, etc. These terms are normally broadly employed in folk psychology to understand, explain, and predict the behaviors of intelligence agents (humans and animals, and potentially artificial intelligences) (Dennett, 1971). These agents are entities to which we attribute the quality of having a mind. Dennett explains that anything whose behavior can usefully and broadly be predicted by taking an intentional stance towards it, is an *intentional system*. The intentional stance is thus a method for interpreting the behavior of an agent by acting *as if* it were a rational agent whose actions are determined by its ‘beliefs’, and ‘desires’ (knowledge and goals, respectively, in parlance of Newell) (Dennett, 2009). An agent’s knowledge of its environment places certain logical constraints (or “boundary conditions”) on its behavior with respect to its achieving its desires (goals). These logical constraints produce patterns in behavior that outside observers (other agents) can pick up on, interpret, and use to predict the behavior of other agents in a given situation. The observer agents then use these interpretations (new knowledge) to determine their own conduct to maximize their chances of attaining their own goals.

10.2 The Context-dependent Nature of the Intentional Stance

Similar to the Pragmatists’ insistence on the context dependency of truth/reality, Dennett argued that intentional attributions were relative to an intentional system’s

environment. Bechtel notes how this aspect of intentional attributions (mental states) being relative to the agent's environment is "somewhat counter-intuitive". Bechtel notes, "What makes intentional idioms seem problematic is that proposition towards which the attitude is held may in fact be false" (Bechtel, 1985, p. 473). Newell and Popper—as well as many others discussing decision-making—focus on the nature of how agents develop internal representations (mental states/knowledge/beliefs) of their environment. In fact, the majority of representations employed in decision-making are dependent on their relationship to the agent's environment because they are representations *of* that environment—that is, the value of these representations lies in their ability to accurately mirror relevant details of the environment they portray.

Predating Bechtel's analysis of Dennett, Newell argued that systems levels such as the knowledge level describe the dynamic relations between agents and their environments, but they do not describe the environment itself (Newell, 1980, p.12). That said, these representations do become part of the intentional system/agent itself because they make internal changes to the agent. As the agent recognizes changes made to the environment—by itself or by another force—it will respond by making changes to its own internal state to keep pace with and account for the changes in the environment. Thus, we see that there is a dynamic relationship between the agent and its environment and the two, in principle, become a system within themselves. Dennett explains, "There is a familiar way of alluding to this tight relationship that can exist between the organization of a system and its environment; you say that the organism continually mirrors the environment, or that there is a representation of the

environment in—or implicit in—the organization of the system” (Dennett, 1987, p.31). It would seem easy to envision an evolutionary narrative of this element of Dennett’s and Newell’s theories that emphasizes how the survival of a species is directly linked to its ability to construct internal representations that accurately capture all the relevant details of that species’ environment. Dennett notes, “What makes some internal feature of a thing a representation could only be its role in regulating the behavior of an intentional system” (Dennett, 1987, p.32). As the agent (or species) is able to construct increasingly accurate internal representations of its environment, it can plan its behavior in ways that increase its chances of survival within an environment—both physical and social environments. The knowledge level is the space in which these internal representations are combined and modified.

This issue of the agent’s mental states being relative to the facts of its environment is similar to the previous discussion of boundary conditions and Pragmatism’s determination of truth existing only within a specific context. Consequently, Bechtel argued that observers could adopt a realist perspective towards intentional attributions of an intentional system—but only within the conditions of the intentional system’s environment. (Bechtel, 1985, p.473). He argues that “what makes intentional idioms seem problematic is that the proposition towards which the attitude is held may in fact be false (Bechtel, 1985, p.437). Philosophers such as Rosenberg (1980) have cited this as a roadblock to treating psychology like a science because two agents might reach different conclusions regarding the same problem/situation—calling into question the objectivity of their individual intentional idioms. However, I think this

is only a lack of objectivity on the surface that provides the illusion of being subjective. When we dig deeper, we will likely learn that the two agents have slightly differing goals and or knowledge of the problem set—which means that they are not, in fact, dealing with the same problem/situation. Variations in either (or both) of these factors—goals or knowledge—would mean the two agents hold different intentional idioms about different situations, which would not call into question the objectivity of either intentional idiom on its own.

To counter this criticism, Dennett's theory provides a way to think about how the design of a complex intelligent system would permit the behavior of the system to be described in terms of its intentional attributions (beliefs), that is, it can be viewed as if it held such intentional attributions (Dennett, 1987)¹. Dennett appears to echo William James' counter to Immanuel Kant's view that objects of belief were not properly objects of knowledge at all and that only objects consisting of sense-content (i.e., physical objects with measurable properties) could count as knowledge². Contrasting Kant's point, James wrote in his *Varieties of Religious Experience*:

"Yet strangely enough they [beliefs] have a definite meaning in our practice....we can act as if there were a God; feel as if we were free; consider Nature as if she were full of special designs; lay plans as if we were immortal; and we find that these words do make a genuine difference in our moral life...Our faith that these unintelligible objects actually exist proves thus to be a full equivalent in praktischer Hinsicht³, as Kant calls it, or from the point of view of our action, for

¹ This further highlights the importance of recognizing the difficulty inherit in the cognitive sciences, which involve a highly self-referential toolset because the object and instrument of study are one and the same. It is this self-reflective nature that makes the study of the mind significantly more difficult than the physical sciences.

² (Immanuel Kant, *Critique of Judgement*, sections 67-77.

³ Valuable practical support.

knowledge of what they might be, in case we were permitted positively to conceive them" (James, 1929, p. 13).

James point "...from the point of view of our action..." is a critical aspect of the discussion, because the physical sciences only describe phenomena in our world, they do not offer insight into what to do about those phenomena as we go about our daily lives. Intentional statements involve value assignments which create bridges from description of phenomena to actions that deal with them.

The primary epistemological claim of the Intentional Systems Theory is that we regularly treat one another *as if* we were intentional systems—i.e., agents with specific intentions/goals and the ability to apply knowledge in a rational manner to attain those goals. Dennett argued that when dealing with intentional systems, if we have the awareness of the constraints placed upon the agent (knowledge of its environment and its desires/goals), then we can do a fair job of predicting the actions that the agent will likely take (i.e., to achieve its desires/goals) in a given situation listing them from best to worst (Dennett, 2009). Although we may not be able to determine what the optimal move is for an agent, we can usually very quickly eliminate all but the most likely actions (i.e., we can *prune the problem space*) in order to better predict the agent's behavior. With this in mind, it is easy to see how Dennett's Intentional Systems Theory effectively deals with the same set of problems about which Newell and Popper theorized.

10.3 Dennett's Perspective on Levels

Dennett employs his own unique system of levels when explaining his Intentional Systems Theory. Dennett notes that there are roughly three "stances" of increasing

complexity (i.e., hierachal levels) which we can employ to predict to make predictions involving various types of objects (Dennett, 2009). The first is the *physical stance*, which Dennett describes as the laborious process of the physical sciences—i.e., it involves whatever we know about the laws of physics and the physical make-up of a given object. This stance serves us well when making predictions regarding less sophisticated objects that are neither alive nor artifacts (designed objects). The second stance that Dennett lays out is the *design stance*, whereby we judge the most probable actions of an object such as an artifact by our knowledge of what it was designed to do. This stance assumes that the artifact was well-designed—i.e., that it will accomplish what it was intended to do. It saves us a great deal of time to analyze artifacts in this manner instead of the longer multi-step methods required by the physical stance. Curiously, although Dennett does not directly address it, this stance also involves an element of intention by virtue of the fact that we attribute some purposes (intention) that some agent had in mind when originally designing the object—that is, it assumes that we correctly assess the object’s original intended purpose. The third level was the *intentional stance*—mentioned previously—whereby the observer uses its knowledge of the intentional system’s beliefs and goals to predict how the system will act in a given environment (context).

10.4 Where does Dennett’s Theory Fall on the Epistemic-Ontic Scale?

Dennett’s position on the epistemic-ontic scale is not an easy one to pin down. He seems to go further than Popper, and yet it is clear that he stops short of formally committing his theory to the full ontological status to which Newell commits his—at

least for the time being and with the current knowledge available. On the one hand, Dennett denies that the intentional stance corresponds to real entities and events (i.e., is not what most could call ontological) (Dennett, 1987). On the other hand, he appears to press for something more than a purely epistemic perspective because he argues that the practical utility of the intentional stance justifies its continued use (Dennett, 1987). Dennett's more conservative epistemological perspective limits the utility of the Intentional Systems Theory. As Bechtel points out in his analysis of the Intentional Stance, Dennett's theory would have greater application to science if it described actual states in the world, as Newell claimed his Knowledge Level theory did (Bechtel, 1985). But perhaps Dennett is merely being overly cautious until neuroscience can provide more insight.

Dennett argues that patterns in human behavior—driven by beliefs—that are revealed by the intentional stance are objective (Dennett, 1987, p.24). Herein arises the confusion over whether Dennett's overall position is epistemological or ontological. If intentional ascriptions (beliefs/knowledge) are objective and the intentional stance reliably allows us to accurately predict how these intentional ascriptions determine behavior, then how does the intentional stance not exist in some way? Put another way, how can the act of holding an objective belief in relation to a physical object be real (objective), but the mechanism that produces this belief not be real? The commonly held perspective seems to hold that objects are real, but relations are not. This, however, seems to come up short because it is often from the complex relations between individual physical objects that functional mechanisms arise.

Dennett views belief/knowledge as objective, while seeing our ability to identify certain beliefs in others as a highly useful predictive strategy but one that has specific boundary conditions—i.e., one that is relative to a specific point of view at the time and place to which the belief in question is linked. Dennett defines an intentional system largely in terms of its relation to the intentional stance, “a system whose behavior is reliably and voluminously predictable via the intentional strategy” (Dennett, 1987, p.15, p.29).

Dennett appears unwilling to commit his theory to a full ontology (as Newell did with his theory) when he says, “while belief is a perfectly objective phenomena...it can be discerned only from the point of view of one who adopts a certain predictive strategy” (Dennett, 1987, p.15). He does, however, appear to consider it to be more than a pure epistemology because he claimed, “the intentional stance description yields an objective, real pattern in the world” (Dennett, 1987, p.34). If the patterns the intentional stance yields are real, it seems difficult to understand how the stance itself could not be because something real does not typically result from something which is not.

Shedding light on where he stands on the epistemic-ontic scale, Dennett argues that we will likely *eventually* be able to map the intentional strategy to an account of how the underlying physical mechanisms work. He explained, “the account of how the strategy [the intentional stance] works and the account of how the mechanism works will (roughly) coincide: for each predictively attributable belief, there will be a

functionally salient internal state of the machinery, decomposable into parts—that is, words or terms” (Dennett, 1987, p. 34). However, it is clear that he does not think we are there yet.

Seemingly coming closer to the position that Newell outlines in his Physical Symbol System Hypothesis, Dennett predicts that someday, “The inferences that we attribute to rational creatures will be mirrored by the physical, causal processes in the hardware; the logical form of the propositions believed to be copied in the structural form of the states in correspondence with them...our brains will eventually be understood as symbol manipulating systems” (Dennett, 1987, p.34). In light of this, it would appear that Dennett believes that we cannot currently prove the objective nature of the intentional stance itself, however, the future will likely reveal that it can be justified by its being mapped to physical mechanisms in the brain once our understanding sufficiently progresses.

Dennett outlines the two key claims made by his Intentional Systems Theory. The first is that “the intentional stance yields descriptions of an objective, real pattern in the world”, which Dennett argues is beyond skepticism (Dennett, 1987, p.34). The second claim is that “this real pattern is *produced by* another real pattern roughly isomorphic to it within the brains of intelligent creatures” (Dennett, 1987, p.34). Dennett explained the second claim was the hypothesis that there is a *language of thought* coded in our brains and that the brain would eventually be understood as a symbol manipulating system analogous with computers. Dennett believed that the first

claim was clearly true, and he assessed that “there are reasons for believing in the second pattern [claim]” but he assessed that the case for the first was much stronger. Instead of dismissing the second claims, Dennett felt that there was enough evidence to support our continued exploration of it and he recommended that “as a matter of scientific tactics that we pursue the hypothesis in its various forms as far as we can” (Dennett, 1987, p.35).

10.5 Similarities between the Intentional Stance & the Knowledge Level

More so than with the theories of Popper and others, there seems to be a co-evolution in the same direction between the research of Newell and Dennett. Newell’s Physical Symbol System Hypothesis appears to have influenced later developments in Dennett’s articulation of his Intentional Systems theory. The Knowledge Level theory appears to be a more philosophically committed version of Dennett’s intentional stance, with Dennett being more cautious regarding what position we can justify, given our current knowledge of neuroscience.

Dennett shares many of Newell’s views on the applications of the mechanism that Newell characterizes as the knowledge level. However, while Newell appeared to view the knowledge level as a real object, Dennett describes his intentional stance in terms of a focus on the perspective of an observer. That is, as we move from the knowledge level to the intentional stance the focus shifts from the phenomenon of the level itself (as an object) to the role of the observer who leverages the knowledge level to make predictions (a perspective) (Newell, 1982; Dennett, 1978).

A closer look the discussions of Newell and Dennett reveals an almost one-to-one mapping of the main terms/components in their respective systems:

Mapping of terminology between Newell and Dennett's theories

Table 7

<u>NEWELL</u>	<u>DENNETT</u>
AGENT	INTENTIONAL SYSTEM
KNOWLEDGE	BELIEFS
GOALS	DESIRE/DESIGN PURPOSE
ENVIRONMENT	SETTING/DOMAIN
PRINCIPLE OF RATIONALITY	RATIONALES FOR ACTION

* Both Newell and Dennett view the context constraints placed on an agent/system's behavior as the result of a combination of the agent/system's environment/setting and goals/desires.

Both Dennett and Newell believed that we do not need to be able to understand the fine-grained internal workings of an agent (e.g., its neural functioning) to be able to predict its behavior; as long as we understand the agent's goals, the environment, and what the agent knows about its environment (Dennett, 2009, pp.4-5; Newell, 1982, pp. 24-25). Dennett writes, "When the information systems (or control systems) are well-designed, the rationales for their actions will be readily discernible, and highly predictive. We needn't know anything about computer-programming to predict the behavior of a system; what we need to know about is the rational demands [constraints based on purpose/goal and environment] of running an oil refinery" (Dennett, 1971, p. 6). By "rational demands" Dennett appears to mean that the constraints set by the combination of the refinery's purpose/goal (to refine crude oil into petroleum) and its environment. This argument is very similar to Newell's discussion regarding how an

observer can predict an agent's behavior if the former understands the latter's goals and knowledge of its environment (Newell, 1982, pp. 20-23). Dennett highlights that we are simply "finessing our ignorance of the details of the processes going on in each other's skulls" and instead simply relying on the practical and well-observed fact that people are, for the most part, fairly rational in their ability to make decisions (Dennett, 2009, p.5).

10.6 Similarities between Dennett's Views and Pattern Realism

Dennett claims that the intentional stance provides advantage points that allow us to discern useful patterns in behavior (Dennett, 1987, p.39). Dennett said, "my view is a sort of realism, since I maintain that the patterns the Martians [observing human behavior] miss are really, objectively there to be noticed or overlooked" (Dennett, 1987, p.37). Dennett explains that a pattern which allows us to make predictions is 'real' so long as it endures and though the pattern may owe its existence to the intentions of an agent, its reality in any interesting sense is independent of the historical facts of its origin (Dennett, 1987, p.39). Consequently, he claims that the Intentional Stance provides a method for discerning similarly useful patterns in behavior and that these patterns are 'real' and objective. He explains, "They [the patterns] are *there* to be detected—but from our point of view they are not *out there* entirely independent of us, since they are patterns composed partly of our own 'subjective' reactions to what is out there; they are the patterns made to order for our own narcissistic concerns" (Dennett, 1987, p.39).

Dennett notes how Anscombe wrote darkly of “an order which is there wherever actions are done with intentions (Anscombe, 1957, p. 80), though she failed to note where this order resided” (Dennett, 1987, p. 39). Dennett himself admitted that it is difficult to say with any certainty where in the intentional order is in the world, “If you ‘look at’ the world in the right way, the patterns are obvious [however]...if you look at (or describe) the world in any other way, they are, in general, invisible” (Dennett, 1987, p.39). It was perhaps this vulnerability of these patterns to shifting perspectives that made Dennett hesitant to conclude that the intentional stance itself had full ontological status.

CONCLUSION:

In closing, Newell asserted that the knowledge level is an actual cognitive mechanism that intelligent agents employ to pick out patterns of behavior in other agents and determine their own behavior in light of how these patterns impact the attainment of their goals. These patterns of behavior are real because they endure, are observable and interpretable across individuals, and support reliable assessments/predictions. The reality of these patterns is a key aspect tying Newell’s theory to others like Popper and Dennett who described this same cognitive mechanism for predicting behavior. Newell’s theory provides a framework for reducing the complexity of analyzing every mechanism involved in goal-directed decision-making by abstracting away details irrelevant to the prediction of intelligent behavior.

The assertion that this mechanism is real is supported by principles of early Pragmatism. It is important to note that Newell was not concerned with determining ultimate objective truth but was instead focused on understanding the role belief played in guiding the actions of agents. Newell's version of what is real is Pragmatic in spirit because it was context-dependent and justified by the reliably predictable patterns of meaning (intentions) that can practically be shown to determine agent behavior. Because these patterns drive agent actions that continually reshape the physical world, they must themselves have some sense of reality. Newell's work, in turn, embodies the Pragmatic principles in computational models that reveal how the brain can be viewed as a Pragmatic engine—i.e. that our brains appear wired for Pragmatism.

Although Newell's version of realism is an uncommon one, it nevertheless provides practical value to cognitive science and artificial intelligence research because it allows us to study and test theories about cognition with increased precision by constructing computational models and computer simulations of the mechanisms that interpret and guide behavior of agents. Consequently, the theory offers valuable insight into current challenges involving agency and knowledge representation in cognitive science and artificial intelligence research.

In addition to improving methods for testing specific theories on cognition, Newell's research can also bring the power of cognitive modeling to a wider audience interested in behavioral analysis and predictive analytics, which rely on a clear understanding of the use of knowledge to attain goals. From here, I hope to further

develop Newell's theory into a framework for more explicit study of the dynamic role that values (meta-goals) play in influencing how agents employ knowledge to accomplish goals, as well as how goals are de-conflicted and re-prioritized as new knowledge becomes available through experience. Doing so could provide insight into how manipulation of specific knowledge (beliefs) and/or goals of agents supports influence, deception, and propaganda efforts.

REFERENCES:

- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- Anscombe, E. (1957). Intention. Oxford, UK: Blackwell.
- Aylesworth, G. (2015). Postmodernism. In *Stanford Encyclopedia of Philosophy* [online]. (<https://plato.stanford.edu/entries/postmodernism/>), Initially published on 30 September 2005, last updated on 5 February 2015). Stanford, CA: Stanford University.
- Ayer, A. J. (1968). *The Origins of Pragmatism*. London, UK: Palgrave Macmillan Publishers.
- Barker, H. (2008). Entry on Charles Sanders Peirce. In N. Chalton (Ed.) *Philosophers: Extraordinary People History*. New York, NY: Metro Books Publishers.
- Bechtel, W. (1985). Realism, Instrumentalism, and the Intentional Stance. *Cognitive Science*. 9, 473-497.
- Bechtel, W. & Hamilton, A. (2007). Reduction, Integration, and the Unity of Science: Natural, Behavioral, and Social Sciences and the Humanities. T. Kuipers (Ed.), From *Philosophy of Science: Focal Issues* (Volume 1 of the *Handbook of the Philosophy of Science*). New York, NY: Elsevier.
- Boole, G. (2003). *The laws of thought*. Amherst, NY: Prometheus Books. (Original work published 1854)
- Brook, A. (2009). The possibility of a cognitive architecture. In D. Denrick & L. Trick (Eds.) *Computation, Cognition, and Pylyshyn*. Cambridge, MA: MIT Press.
- Causey, R. (1977). *Unity of Science*. Dordrecht, Holland: Reidel Publishing Company.
- Chakravartty, A. (2017). Scientific Realism. In *Stanford Encyclopedia of Philosophy* [online]. (<https://plato.stanford.edu/entries/scientific-realism/>), Initially published on 27 April 2011, last updated on 12 June 2017). Stanford, CA: Stanford University Online.
- Chalmers, D. (1996). *The Conscious Mind*. London, UK: Oxford University Press.
- Chemero, A. (2009). *Radical embodied cognitive science*. Cambridge, MA: MIT Press.
- Chomsky, N. (2016). *What Kind of Creatures are We?* New York, NY: Columbia University Press.

- Churchland, P. S. (1986). *Neurophilosophy: Toward a Unified Science of Mind-Brain*. Cambridge, MA: MIT Press/Bradford Books.
- Churchland, P. S. (1987). Epistemology in the age of neuroscience. *Journal of Philosophy*, 84: 544-53.
- Churchland, P. (1994). Betty Crocker's Theory (a review of *The Rediscovery of the Mind* by John Searle. *London Review of Books*, 12 May 1993: pp.13-14.
- Churchland, P, Kock, C., & Sejnowski, T.J. (1990). What is computational neuroscience? In E. L. Schwartz (Ed.), *Computational Neuroscience* (pp. 46-55). Cambridge, MA: MIT Press.
- Dawson, M. R. W. (2013). *Mind, body, world: Foundations of cognitive science*. Edmonton, AB: AU Press.
- Dennett, D. C. (1971). Intentional Systems. *Journal of Psychology*, LXVIII, 87-106.
- Dennett, D. C. (1987). *The Intentional Stance*. Cambridge, Massachusetts: The MIT Press.
- Dennett, D. C. (1991). *Consciousness explained*. Boston, MA: Little, Brown.
- Dennett, D. C. (2009). Intentional Systems Theory, In B. McLaughlin, A. Beckermann, & S. Walter (Eds.), *The Oxford Handbook of Philosophy of Mind* (pp. 339-50). London, UK: Oxford University Press.
- Descartes, R. (1960). Discourse on method and meditations. (L.J. LaFleur, trans.). Indianapolis, IN: Bobbs-Merrill. (Original work published 1637)
- Fodor, J. (1974). Special Sciences (or: the disunity of science as a working hypothesis. *Synthesis*, 28, 97-115.
- Fodor, J. (2010). *LOT 2: The Language of Thought Revisited*. Oxford, UK: Oxford University Press.
- Hofstadter, D. (1999). *Gödel, Escher, Bach: An Eternal Golden Braid*. New York, NY: Basic Books
- Honderich, T. (2005) Empiricism Entry. In T. Honderich (Ed.) *The Oxford Guide to Philosophy*. New York, NY: Oxford University Press.
- Humbolt, W. (1988). On Language: On the Diversity of Human Language Construction and Its Influence on the Mental Development of the Human Species. (Peter Heath, trans.). New York: Cambridge University Press.
- Hunt, E. & Luce, R. D. (1992). SOAR as a world view, not a theory. *Behavior and Brain Sciences*, 15, 425-449.

James, W. (1929). *Varieties of Religious Experience: A Study in Human Nature*. New York, NY: Random House Publishers. (Originally published in 1920)

James, W. (1981). *Pragmatism: A New Name for Some Old Ways of Thinking*. Indianapolis, IN: Hackett Publishing. (Original work published in 1907)

James, W. (1996). *Essays in Radical Empiricism*. University of Nebraska Press. (Original work published in 1912)

James, W. (2000). *Pragmatism and Other Writings*. New York, NY: Penguin Books. (Original work published in 1907)

Kant, I. (1996). *Anthropology from a Pragmatic Point of View*. Dowdell, Victor L. (trans.) Southern Illinois University Press. (Original work published in 1798)

Kersten, L., West, R., & Brook, A. (2017). Leveling the Field: Talking levels in cognitive science. *Cognitive Science*. 9, 473-497.

Koffka, K. (1935). *Principles of Gestalt Psychology*, London, UK: Lund Humphries.

Laird, J., Newell, A., & Rosenbloom, P. (1987). SOAR: An architecture for general intelligence. *Artificial Intelligence*, vol. 33, 1-64.

Lieto, A., Lebiere, C., & Oltramari, A. (2017). The Knowledge Level in Cognitive Architectures: Current Limitations and Possible Developments. *Cognitive Systems Research*. June 2017, 1-42.

Locke, J. (1999). An Essay in Human Understanding (abridged). In R. Cummins & D. Owens (Eds.) *Central Readings in the History of Modern Philosophy*. Belmont, CA: Wadsworth Publishing Company. (Locke's Essay was originally published in 1690)

Marr, D. (1982). *Vision*. San Francisco, CA: W.H. Freeman.

Marr, D., Poggio, T. (1977). From understanding computation to understanding neural circuitry. *Neurosciences Research Bulletin*, 15, 470-488.

Newell, A. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice Hall.

Newell, A. (1973). You can't play 20 questions with nature and win: projective comments on paper of this symposium. Carnegie Melon University, Department of Computer Science Research Showcase series. Pittsburgh, PA.

Newell, A. (1980). Physical symbol systems. *Cognitive Science*, 4, 135-183.

- Newell, A. (1982). *The Knowledge Level*. Pittsburgh, PA: Carnegie Mellon University, Department of Computer Science Research Showcase series.
- Newell, A. (1990). *Theories of Unified Cognition*. Cambridge, MA: Harvard University Press.
- Newell, A., & Simon, H. (1976). Computer Science as Empirical Inquiry: Symbols and Search. In J. Haugeland (Ed.), *Mind Design II: Philosophy, Psychology, and Artificial Intelligence* (1997). Cambridge, MA: The MIT Press.
- Nickles, T. (1973). Two concepts of intertheoretic reduction. *The Journal of Philosophy*, 70, 181-201.
- Peirce, C. S. (1978). *The Philosophy of Charles Sanders Peirce: Selected writings*. In Buchler, J. (Ed). New York, NY: AMS Press.
- Peirce, C. S. (2016). *Philosophical Writings of Peirce*. In Buchler, J. (ed). New York, NY: Dover Publications. (Originally published in 1940)
- Peterson, J. (1999). *Maps of Meaning*. London, UK: Routledge Press.
- Popper, K. (1960). *The Poverty of Historicism*. New York, NY: Routledge Publishers.
- Popper, K. (1994). *The Myth of the Framework*. New York, NY: Routledge Publishers.
- Popper, K. (2002). *The Logic of Scientific Discovery*. New York, NY: Routledge Publishers. (Originally Published in 1959)
- Pylshyn, Z. W. (1984). *Computation and Cognition: Toward a Foundation of Cognitive Science*. Cambridge, MA: MIT Press.
- Pylshyn, Z. W. (2003). Return of the mental image: Are there really pictures in the brain? *Trends in Cognitive Sciences*, 7(3), 113-118.
- Pylshyn, Z. W. (2003b). *Seeing and visualizing: It's not what you think*. Cambridge, MA: MIT Press.
- Rosenberg, A. (1980). *Sociobiology and the Preemption of Social Science*. Baltimore, MD: Johns Hopkins University Press.
- Schiller, F.C.S. (1912). *Humanism: Philosophical Essays*. London, UK: Macmillan and Co. LTD. (Originally published in 1903)
- Schiller, F. C.S. (2005). *Studies in Humanism*. Boston, MA: Adamant Media Corporation-Elibron Classics Series. (First published in 1907)

Scribner, S. & Tobach, E. (1997). Mind and social practice: *Selected writings of Sylvia Scribner*. Cambridge, UK; New York, NY: Cambridge Press.

Simon, H. (1969). *The Sciences of the Artificial*. Cambridge, MA: MIT Press.

Suchman, L.A. (1987). *Plans and situated action: The problem of human-machine communication*. New York: Cambridge University Press.

Thagard, P. (2009). Why philosophy needs cognitive science and vice versa. *Topics in Cognitive Science*, (1) Cognitive Science Society, 237-254.

Vera, A. H. & Simon, H. A. (1993). Situated action: A symbolic interpretation. *Cognitive Science*, 17, 7-48.

West. R. L. & Young, J. T. (2017) Proposal to add Emotion to the Standard Model. Association for the Advancement of Artificial Intelligence (AAAI) 2017 Conference Proceedings.

Wimsatt, W. C. (1976). Reductionism, levels of organization, and the mind-body problem. In I. Savodnik (Ed.), *Consciousness and the Brain: A Scientific and Philosophical Inquiry*. New York, NY: Plenum Press.

Winograd, T. & Flores, F. (1987). *Understanding computers and cognition*. New York, NY: Addison-Wesley.