12-2015

Disentangling Embodied Cognition: An Examination of the State, Problems, and Possibilities of Embodied Cognition

Cody Cash
University of Arkansas, Fayetteville

Disentangling Embodied Cognition:
An Examination of the State, Problems, and Possibilities of Embodied Cognition

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Philosophy

by

Cody Cash
Missouri State University
Bachelor of Arts in Philosophy, 2000
University of Arkansas
Master of Arts in Philosophy, 2010

December 2015
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

____
Dr. Jack Lyons
Dissertation Director

____
Dr. Edward Minar
Committee Member

____
Dr. Eric Funkhouser
Committee Member

Abstract

Embodied cognition has received a fair amount of attention in philosophical, neuroscientific, and robotic research during the past several decades, yet the precise nature of its goals, methods, and claims are unclear. This dissertation will ascertain and examine the primary themes in the field of embodied cognition as well as why, and if, they offer significant challenges to traditional cognitive science models. Though many theories believe they are providing accounts that should replace traditional models, to do so they will have to overcome the very difficult challenge of arguing that mental content and capabilities derived from sensorimotor activity can continue to function independent from the sensorimotor processes necessary for their instantiation. In short, they have to rule out brain-in-a-vat scenarios. Upon examination, most embodiment theorists either do not attempt to address this fundamental issue or they fail to provide a successful account of how it can be achieved through a model of embodied cognition. And for those who are able to overcome this obstacle, doing so will require reconsidering the brain/mind as an extension of the body once it has become thoroughly entangled with sensorimotor processes and activities.

The first portion will attempt to clarify the various claims of embodied cognition to understand what the theories are saying and how they are at odds with traditional cognitive science. Then, several of the most significant embodied arguments will be looked at in detail to determine their philosophical and empirical strength. Finally, the fundamental issue of the metaphysical dependence of mental content and capabilities on embodied activity will be laid bare and it will be shown that nearly all embodiment theories are in alignment with traditional cognitive science on this topic, and those that aren't will require some additional work.

Acknowledgments

Special thanks to Dr. Jack Lyons for his guidance and assistance throughout the writing of this dissertation. His knowledge and eye for underlying issues were crucial to the completion of this work.

Dedication

This dissertation is dedicated to my wife, the lovely Allison. Her unwavering encouragement is something I always cherish, and the main reason this work was ever undertaken and finished.

Table of Contents

1. Introduction

No one exists for even an instant without performing action
-Bhagavad Gita

The term "embodied" cognition is both wonderfully clear and frustratingly vague. In regards to clarity, at least it rules out any notion of an immaterial mind. The mind and cognition are, at minimum, constituted by or embodied via some physical components. There is no immaterial Cartesian mind and thus no ghost in the machine magically pushing and pulling levers or selectively exciting neurons; the mind is, and can only be, in the flesh, so to speak. Due to this, any embodied theory of cognition rules out substance dualism and does not consider the mind and body as wholly separable entities. And by "body" these theories mean much more than the brain. Though the brain plays an indisputably integral role in cognition, embodied theories generally take the brain to be one aspect of the body essential for cognitive activity—just as the eyes, ears, tongue, toes, fingers, and elbows (or some minimal combination of physical components) may also be essential. This does not mean that we must consider the entire body at all times when studying or discussing the mind, but it does entail that our cognitive capacities do not exist or operate fully autonomously from the body; the mind's activities and/or content are tied to bodily engagement with the world and one's ability to perceive that engagement (sensorimotor activity). This point signals a departure from classical cognitive science and folk psychology.

Traditionally, from casual conversations between parents and children to depictions in all forms of media, "reason" has been portrayed as a unique faculty of the human mind that is independent of our bodily movements and senses. We may go out and engage with the world in deliberate ways—finding food, walking, laughing with others, playing sports, etc.—but we can also retreat to an inner world of thought that seems to be able to function separately from our body. For example, it's possible to walk fluidly among a group of people on a crowded sidewalk without bumping into anyone or taking a wayward step, while at the same time contemplating the ingredients you'll need to get at the grocery store to make chocolate chip cookies. In such instances the body and mind are both active, yet they seem to be operating independently. This is at least one important aspect of what is believed to separate us from other animals like squirrels, lobsters, and house pets—they live more reflexively in the moment, whereas we humans can extract our minds from the act itself to a higher level of thought, considering the act from a different perspective or contemplating something else entirely. Yet, even in the face of this simple, everyday example that many take as folk evidence of the mind's independence from the body, embodiment theorists say such conclusions about the mind and cognition are utterly false. If we take a deeper look at such scenarios, they say, we will find clear evidence that all cognition—even that which occurs in moments of inaction or removed from somatic activity—is a combination of activity between mind, body, and world, rather than a purely neural or magically immaterial event.

Saying bodily engagement with the world affects the mind is not a new or surprising point. The distinctive aspect of any embodied theory of cognition is *how* mind, body, and world are tied together. And this is where the term "embodiment" can become frustratingly vague. At minimum, most theories will say mental content is based on the unique interaction we have with the world through the types of bodies we have. Thus, the fact that humans have visual and auditory receptors able to receive and process input in a fairly specific and limited range of frequencies and wavelengths entails that any concepts we have cannot exist outside of that range. Take infrared light: though one can have a concept of infrared light, one can only perceive and understand it via colors that fall within the human optical window from violet to red. This is why night vision equipment is (relatively) bulky and depends upon advanced engineering. Such devices rely upon complex gadgetry to first convert ambient light photons into electrons, then amplify them via chemical and electrical processes, and then convert the electrons back into visible light. What was previously invisible—because human eyes could not directly perceive the wavelengths—now appears in shades of green that one can clearly and easily process. Though one might put on night vision goggles and believe s/he is experiencing infrared light, s/he is still only experiencing the same optical window as usual but now that optical window is acting as a translation of sorts for light input that actually falls outside of it. In this way, the colors that supply content to our concepts are wholly dependent upon the perceptual systems our bodies have. Every mental concept or mental representation a human being has will have a color that exists at some point within the range of violet to red that human eyes can perceive. This is why we don't have concepts of infrared apples or gamma ray stoplights, and why human concepts of X-rays are actually shades of grey, black, and white—we are conceptually limited by the input our bodies are able to perceive. If the eye's abilities or function were slightly different, then one's perception of colors could be different, which entails that any conceptual content one has involving color(s) would be different as well. On its own, this should not seem revolutionary. The claim that our conceptual content is dependent upon our embodied engagement with the world is fairly minimal in that it alone does not offer a significant shift in cognitive science or philosophy of mind.

On the more extreme end of embodied cognition, some theories suggest that not only are the contents of our concepts based on bodily engagement, there are also no such things as symbolic computational processes in the brain at all. From the example in the previous paragraph, one could accept that the content of mental representations comes from engagement with the world. The concept STOP SIGN, for instance, has the shape and color information it has based on one's visual perceptual system, and the color, shape, and size of the sign are represented via formal mental symbols that are both encoded and decoded during the cognitive process. In such a process, one can consider the brain as a central processing unit taking input from the outside and then translating it into mental symbols that can be cognitively manipulated (I'll discuss this model more in the next chapter). According to traditional cognitive science, the central processing unit is where thought takes place. Sure, it won't have much to think about without input, but just as a computer is still a computer even when it's not turned on, a mind is a mind as long as it can process information, and the brain alone is both what gives and retains this potential. According to embodied cognition, it is incorrect to consider the brain by itself as the mind once removed from all its input devices and perceptual systems. But, this alone still allows for most of the traditional cognitive model to stay intact. There can still be a formal, rule-governed language of thought for the most part, but that language is now inextricably tied to the body. Representationalism and computationalism can still be largely true, just with an embodiment amendment of sorts. However, theories that fall on the more extreme end of the embodied spectrum claim the representational and computational theories of mind are false. Theories that go this more radical route support the notion that cognition is constituted by a unique sort of interaction between mind, body, and world. And some even go so far as to take a strict noncomputational and nonrepresentational view that relies heavily on dynamical systems theory to help explain how cognition can operate without any sort of symbolic mental representations.

There are assortments of implications that accompany the broad-ranging theories that fall under the ambiguous umbrella of "embodiment." At one end the theories are very much in line with traditional cognitive science but are attempting to show how the body has a much greater role than previously believed, and on the other end are theories which aim to deny nearly all folk psychological stances on the mind and replace theories that have been both exceptionally influential and largely successful over the past several decades. One could and should look at this range of implications and wonder how theories purported to be under the same heading of embodiment could lead to such diverse and sometimes contradictory conclusions. Quite simply, there is too much ambiguity. The aim of this work is to get rid of, or at least work through, some of this ambiguity and discern where, how, and if embodiment theories offer a significant challenge to the traditional models of cognitive science. With this end in mind it will be helpful to begin with a brief scan of descriptions found in the existing literature on the subject of embodied cognition.

The difficulty of pinning down exactly what is at stake and how the theories operate is perhaps most obvious in resources that aim to provide an objective summary of embodied cognition, such as the Stanford Encyclopedia of Philosophy and the Internet Encyclopedia of Philosophy:

> Embodiment Thesis: Many features of cognition are embodied in that they are deeply dependent upon characteristics of the physical body of an agent, such that the agent's beyond-the-brain body plays a significant causal role, or a physically constitutive role, in that agent's cognitive processing. (Wilson and Foglia, *SEP*, 2011)

> Embodied Cognition…emphasizes the formative role the environment plays in the development of cognitive processes. The general theory contends that

> cognitive processes develop when a tightly coupled system emerges from real-time, goal-directed interactions between organisms and their environment; the nature of these interactions influences the formation and further specifies the nature of the developing cognitive capacities. (Cowart, *IEP*, 2005)

Both of these definitions reveal the central theme—mind is not limited to the brain; cognition is dependent on interaction with the world—but, at best, they *hint* at what is going on rather than making anything explicit. Which features of cognition are embodied? What parts of the body are they dependent upon and how does that dependency operate? Is the role of the body causal or constitutive? And what is meant by "a tightly coupled system?" These are some of the questions that are not clear from the statement of the thesis alone, which might be expected. After all, encyclopedic resources are meant to provide an overview, and an overview can only be as thorough as the resources from which it is pulling. Thus, the blame for this lack of specificity, or inability to provide specifics due to conflicting accounts, is likely due to the work going on in embodied cognition. And this becomes evident when looking at individual theories of embodied cognition. Here are just a few to consider,

> By using the term *embodied* we mean to highlight two points: first, that cognition depends upon the kinds of experience that come from having a body with various sensorimotor capacities, and second, that these individual sensorimotor capacities are themselves embedded in a more encompassing biological, psychological, and cultural context. (Varela, Thompson, and Rosch, 1991: 172-173)

> The notion of an embodied cognition is this: humans can and do use propositional logic to describe and think about their experiences. However, the stuff that our logic works on is nonpropositional and, indeed, is totally based on bodily experience. We deal with our perceptions and actions in terms of fluid, dynamic, contextual categories, patterns of organization, which form the very grist for our engagement of meaning. In an enactive cognition, meaning in the most abstract sense cannot be separated from actions. Meaning has its origins in actions and is made manifest—created—in real time and through activity. (Thelen and Smith, 1994: 323)

> An embodied concept is a neural structure that is actually part of, or makes us of, the sensorimotor system of our brains. Much of conceptual inference is, therefore, sensorimotor inference. (Lakoff and Johnson, 1999: 20)

> Human life and the beginnings of the intelligent behavior that we can see in the infant are not only measured by their physical manifestations as bodily processes, they *are* those processes, and are constituted by them. Movement and the registration of that movement in a developing proprioceptive system (that is, a system that registers its own self-movement) contributes to the self-organizing development of neuronal structures responsible not only for motor action, but for the way we come to be conscious of ourselves, to communicate with others, and to live in the surrounding world. Across the Cartesian divide, movement prefigures the lines of intentionality, gesture formulates the contours of social cognition, and, in both the most general and most specific ways, embodiment shapes the mind. (Gallagher, 2005: 1)

Each of these explanations comes from a philosopher, or philosophers working in conjunction, to provide his/her/their own unique take on embodiment. And the same problems that appeared in the encyclopedic definitions appear in these as well. For Varela, Thompson, and Rosch, sensorimotor capacities are essential. But which ones, how, and are they essential only causally or as a metaphysical constituent of cognition? For Thelen and Smith, meaning is made through activity, but which meanings, which activities, and can meanings be realized through multiple activities or are they bound to specific ones? For Gallagher, intelligent behavior is constituted by bodily processes because a proprioceptive system contributes to the development of neuronal structures, but how does this occur, and do the neuronal structures remain and function even if the brain becomes somehow disembodied? Each of these philosophers attempts to answer these questions. Unfortunately, the answers are not always clear or satisfying. More importantly, the answers are often at odds with one another. For example, two theories may agree that cognition is impossible without sensorimotor capacities, yet one may conclude that these capacities and their activities play causal roles by providing the semantic content of a given mental representation of an object, whereas the other may claim that the object is directly perceived without any mental representation whatsoever.

This range of possibilities underscores the need to examine the existing literature on embodied theories of cognition. Without some clearer understanding of the claims being put forth and the stance(s) to be examined, it won't be clear whether empirical evidence or philosophical reasoning is solid support for embodied theories that are making significant claims against traditional models. Therefore, a large portion of this work will be dedicated to providing an overview and taxonomy of the main theories.

Some taxonomies of embodied cognition theories currently exist, dividing the theories along lines based how the body operates in each theory or how the theories compare to traditional cognitive science. Some involve dividing the theories based on whether they consider the body as a constraint on the nature and contents of the cognitive system, as a distributor of computational and representational loads between neural and non-neural structure, or as a regulator that ensure cognition and action are tightly coordinated (SEP, 2011). Or, alternatively but similarly, dividing them based on whether the theories consider the body and sense organs to determine all concepts in the head, or consider the body and sense organs to constitute cognition itself (thus taking cognition out of the head), or consider all representational models—in or beyond the head—to be hopelessly defective and in need of replacement (Shapiro, 2011). I will take a slightly different route since I am more interested in deciphering those aspects of embodiment theories that offer the most significant and strongest challenges to the traditional model.

In the next chapter, I will provide a brief overview of the traditional model to lay out the key principles embodiment theories take issue with. This overview is not intended to be comprehensive. In fact, it will focus primarily on the claims of Jerry Fodor regarding the Language of Thought hypothesis.

The third chapter will be the taxonomy of embodied theories. Rather than lumping entire theories into a single stance I will instead look at six core, often common themes at the heart of most embodied cognition theories:

1. Mind, body, and world are "coupled"
2. Perception is perceptually-guided action
3. Mental representations (assuming they exist) are not amodal
4. Language is based on embodiment
5. Efforts in artificial intelligence support embodiment

For the most part, these themes and claims appear in many embodiment theories. Some, such as (1), are ubiquitous. However, others, such as (3), are more radical within the embodiment community and are tied to a limited set of theories. This list will certainly be incomplete to some. Several philosophers (Gallagher, Merleau-Ponty, Sartre) also claim the self-awareness that provides the Cartesian cornerstone of Western philosophy is attainable only via embodied cognition. However, I am choosing to focus on those claims that have a more scientifically evaluable stance (although Merleau-Ponty will be discussed in the coupling section). This is not to say the issue of self-awareness and embodiment are separate issues. I tend to agree with those who believe the two are inextricably bound together, but for the purposes of this discussion it would take us too far afield from the studies and neurological findings that will be the focus of this effort. Additionally, I will not cover the issue of extended cognition—which claims that not only is the mind embodied but also, in some instances, aspects and objects of the external environment can become part of one's cognitive system. This too is an extremely intriguing and valuable topic. However, it is supported by very few (Andy Clark being the most prominent) and many embodied theorists are devout in their efforts to distance themselves from such claims. Plus, though cognitive extension seems to entail embodiment, the reverse is not true. Therefore, we can debate embodiment without needing to include extension. In detailing each of these claims, criticisms and potential compromises will be discussed. I will attempt to do justice to the critics but will spend more time addressing the positive arguments than the cases against them in order to provide a clearer picture of exactly what the different embodiment theories are saying.

In the fourth chapter, I will discuss what I believe to be a fundamental dividing line that can be drawn across embodiment theories to determine if they are truly at odds with traditional cognitive science models. Though some pillars of thought in each camp are adamant about the incompatibility of the two approaches, once we examine the underlying metaphysical relationships supported by these theories, not all are fundamentally at odds with traditional cognitive science models. In fact, nearly all aren't. Most end up falling into an area that helps explain how the body is more intertwined with mental representations than previously believed, but not necessarily providing evidence that traditional models need to be wholly replaced.

In the final chapter, I'll argue that part of the reason for the seeming compatibility between traditional and embodied theories discussed in the previous chapter is due to the common way in which we define and describe the brain. However, as an organ uniquely shaped (in terms of modular function and sensorimotor processing) by embodied interaction with the world, the brain could be considered part of the body once it has become entangled with embodied activity. Recasting the brain in this way would definitely call for a replacement of traditional cognitive science models in favor of an embodied approach. But without recasting the brain in this way, nearly all embodiment theories discussed here are not making claims fundamentally at odds with traditional cognitive science. However, there's much ground to cover before reaching that point. Let's get started.

2. A brief overview of the traditional model

An exhaustive presentation of the stances of traditional cognitive science is not necessary in order to understand the ways in which it is at odds with theories of embodied cognition, but a short overview will be helpful. What I will present here is essentially a caricature of arguably the most influential model in cognitive science: the Language of Thought hypothesis proposed by Jerry Fodor. I will present it within the context of the overall traditional endeavor, and will call out unique aspects as needed. This is not to suggest this particular model encompasses all there is to be said about traditional cognitive science models. Just as with embodied theories, there are a great many variations and diverging streams of thought. However, Fodor has been exceptionally influential and has formed the foundation of nearly all approaches that embodied cognitivists argue against. Thus, consider this section as a target for embodied cognition theories. When we get to some of the problems facing embodied theories, it may be that this target is too narrowly construed and successfully overcoming the notions posed here may still not be enough to replace the traditional model with an embodied one. But this will at least provide an overview for what is at the heart of the debate.

The key to the traditional, or standard, cognitive model is symbols. Symbols of any kind act as stand-ins for something else, whether it's a company, a country, a word, a concept, etc. And, as stand-ins, they are not identical to the things for which they stand; otherwise they would be the actual things and not symbols. This is why two golden arches can come to stand for McDonalds, a rectangle with horizontal stripes of red, white, and blue can stand for the Netherlands, "cat" can stand for a small, four-legged furry despot, and the emoticon ☹ can stand for sadness. The thing or concept being signified doesn't need to be present in front of a perceiver if the symbol is working properly. Thus, one can read three letters, such as C-A-T, and think of an animal that exists in the world without needing an actual cat to be in direct view. The value of symbols is that they can transform and translate things into different mediums of communication and perception. An image can mean a word (such as a red octagon meaning "stop"), a word can mean an object in the world (such as the term "octagon" meaning an actual eight-sided shape), an object in the world can mean an abstract concept (such as doves signifying peace), and so on. However, a symbol is only meaningful within a given symbolic system. For example, a circle on its own has no inherent meaning, but when considered within the symbolic system of Arabic numerals it means "zero," or when considered within the symbolic system of the English language it can also be the vowel "o."

Whether dealing with Arabic numerals in the context of algebra, English words in the context of written communication, or even 1s and 0s in binary code, such systems rely on rules. Otherwise, a circle could stand for zero in one algebraic equation and for seventy-three in the next. The rules of a system govern both the initial meaning of a given symbol and how that symbol functions when combined with other symbols in the system. In some systems this is fairly basic, such as with emoticons. One side of a parenthesis is a mouth; colons and semicolons signify eyes; and how the two are put together signifies a range of emotions including happiness, sarcasm, and surprise. However, when considering something like the English language, the rules become much more complex. There is the designation of certain shapes (and combinations of shapes) to mean a letter that has a corresponding sound, then the designation of certain combinations of letters to denote individual word concepts, and the designation of the order in which words can be combined with one another to express thoughts. The second is the semantic content of words and the third is the syntax that governs how words can be related to one another. And in the context of natural language, even though each of these could be considered somewhat arbitrarily designated at the inception of the language (the shape designated the sound and role of “o” could have looked like a “d” or a “3” and still been assigned the same sound and role; or verbs could have been placed at the end of sentences), once the system is adopted and put into practice those arbitrary designations govern when and if the language is functioning correctly. Thus, “lkj87rkb” is meaningless in English while “hello” is meaningful; “Pencil poppy jacket slander” does not effectively communicate a thought (though it sounds curiously like a would-be band), whereas “Frank likes to run on Saturdays” does. Words with semantic content combined in the correct manner, according to the syntax of the language, allow for thoughts to be expressed and understood.[1]

Additionally, because this is all about rule-governed *symbols*, one can analyze and comprehend a symbolic system, like a natural language, without the need to understand the actual objects in the world being represented by the symbols of that language. From the last example, it doesn’t matter who Frank really is in order to understand the sentence. Though there are certain real-world inferences that can be made based on the rules governing the English language—for example, “Frank” is identified as the subject of the sentence, thus implying that Frank is likely a person, place, or thing—the sentence can be comprehended (or at least make

[1] Though rules may be arbitrary in some symbol systems, such as natural language, this is not true of all systems. Traffic signs, for example, utilize simplified images that correspond to identifiable, real-world objects. If the symbols or rules governing traffic sign communication were completely disassociated from the physical objects or events they describe (with, say, diamond shapes representing the curviness of the upcoming road or a series of five dots representing a school zone) the signs would likely fail to function as they do—non-arbitrariness seems essential to such a system because it is intended to communicate effectively to any persons traveling on those roads, regardless of which natural or symbolic language they generally rely upon.

sense minimally) to someone regardless of whether Frank is actually an adult male, a fish, or a chair. The truth or falsity of the sentence will require designating a specific person, place, or thing as Frank and finding out whether he, she, or it likes to run on Saturdays, but the language and the thought it expresses can be understood regardless of how well it describes actual events in the external world.

This is just a basic primer on symbols. But symbols and the particular manners in which they are manipulated are the central concern and cornerstone of traditional cognitive science. According to the traditional account, cognition is believed to operate similar to a natural language in that it relies on the appropriate manipulation of symbols that stand-in for perceptions, thoughts, objects, and ideas. And just as natural languages are based on either actual words on a page as well as perceptible sounds or hand signals communicated among individuals, mental symbols are physical entities too. Exactly what physically constitutes the symbols—specific neurons, clusters of neurons, or patterns spread out over an entire neural network—is debatable between theories and research, but what separates mental symbols from other brain matter is that they are physical *and* have semantic content. In contrast, something like a proton, though physical, is just a proton; it doesn't *mean* anything in addition to its physical properties.

Cognition, then, can basically be viewed as the rule-governed operation and use of mental symbols within the brain. This is interesting, but vague. Questions still remain as to whether all mental symbols are necessarily language-like in ways analogous to the examples already mentioned, how they come to have semantic content, and how they combine to form the incredible range of propositions, feelings, and memories expressed in cognition.

In regard to the language-likeness of symbols, for now we will progress assuming all mental symbols are language-like in that each symbol has some semantic content and the ability to combine with other mental symbols to form further concepts, propositions, physical responses, and so on. This assumption also implies that the ability to combine is governed by a syntax of some sort—a set of rules that delineates which symbols can go together and how they can do so, as well as the types and tokens of mental content resulting from such combinations. With this in mind, if the symbols are based in clusters of neurons and the combinations of such clusters, each cluster must have something meaningful represented within it. Let's say a given cluster represents the smell of a rose. If the cluster lacks this or any other semantic content then no mental symbols exist within or due to that specific cluster of neurons. Yet the representation of the smell alone does not make the cluster of neurons a constituent of a mental concept. On its own it's just a representation of a smell. To become a symbol that can come to represent the smell of a rose it must also possess certain abilities. For one, syntactically, it must be able to play a distinct role in propositional attitudes. The smell is *of* a rose, and not just an indiscernible sensation; it belongs to *that* rose, or to roses in general, thus being able to function as a perceptible property belonging to object concepts. Additionally, the symbol has the semantic ability to combine with other clusters of neurons responsible for, say, the visual image of a flower with red petals, awareness of the letters "R," "O," "S," and "E," or the entire word "Rose." These symbols can then combine to form the concept "ROSE," complete with a visual image, scent, texture, and natural language expression, which will then have rules affecting its ability to combine with other concepts to form propositions about roses or involving roses. Importantly, because mental symbols are physical entities both the semantics and syntax of mental symbols must also be physical in that they are not represented (in the same way as the concept "ROSE"). Instead, they are properties of the representations instantiated by the neuron clusters. More specifically, the semantic content is a property of that specific symbol token, but the syntax is a property of the symbol type as well. Perhaps only that specific symbol represents the smell of a rose but all symbols representing smells will have certain shared combinatorial powers and roles. When a cluster of neurons successfully produces representations having such semantic content and syntactical abilities, then it is a mental symbol that can become a constituent of cognition.

Though brief, from this discussion a standard has been established that all mental symbols are necessarily language-like. Not surprisingly, this criterion is debatable and it will be discussed again in section 3.3 when the topic of non-amodal representations comes up, but it can be considered the standard view.

The other questions (how mental symbols come to have semantic content and how they combine to form the massive range of propositions, feelings, and memories of thought) are easier to answer with this language-likeness in place. Any thought one has is really just an experience or relationship of certain representational mental symbols. Thus, something like the belief "Frank likes to run on Saturdays" is a composition of formal mental symbols that represent the individual concepts "Frank," "like," "run," and "Saturday," that have been combined in the right manner to allow the thinker to understand Frank is a person, place, or thing that enjoys a certain form of exercise on a specific day of the week. And, just as jumbling the words to read "Saturdays run like Frank" changes the meaning of the sentence, if the formal mental symbols are put together differently, different concepts and thoughts will emerge. In this way, mental symbols generate representations that function like words in a language of thought—"Mentalese," as Jerry Fodor called it.

Mentalese is an exceptionally unique and important notion. I have described symbols as "language-like" many times already, but there is a wide variety of forms a language could take.

Mentalese provides an account of thought differentiated from norms of natural languages or road signs and considers the meta-language, which allows for any and all other languages to be possible. As the language behind all other languages, the best we can do in our effort to describe Mentalese is to rely on language analogies. And, though I've used natural language analogies so far, for Fodor that is not the best way to illustrate and understand the language of thought. Instead, he and others in traditional cognitive science describe the manner in which the mind manipulates these symbolic representations and the syntactical properties of the representations that govern their manipulation as a computational language process. This is because the operation of the language of thought and computer programs is believed to function similarly. Input is encoded in a manner that imbues representations with unique semantic content and syntactical capabilities, it is then processed and/or stored, and output follows based on the inherent syntax governing the system. "Computational processes are ones defined over syntactically structured objects; viewed in extension, computations are mappings from symbols to symbols; viewed in intension, they are mappings from symbols under syntactic description to symbols under syntactic description." (Fodor, 1994: 8). Cognition is defined by the manipulation and interaction of symbolic representations similar to, though not necessarily identical to, the language(s) of computer programming.

This can have several unnerving consequences for embodied theorists. One being that it leaves open the possibility that an effective cognitive science need not concern itself with the body's direct interaction with the world (input or output). Instead it can focus on bodily interaction indirectly by studying how neural activity is able to store and process the input and how it is able to generate output. It does not matter if the input and output are coming or going via eyes, cameras, mechanical arms, or electrical stimuli. All that truly matters is what happens in-between, which is the underlying computational process: the language of thought. Focusing on the in-between is what Susan Hurley calls the “classical sandwich.” In this sandwich, cognition exists nestled between input from the external world and output to it. If this sandwich model is accurate, the representations that operate within cognition may be doing so in an amodal manner by functioning in a way that is not tied to the perceptual system(s) that provided the input necessary to generate the semantic content of the representation, nor tied to any motor system(s) that may provide output. Perceptual and motor systems causally affect representations, and may be metaphysically necessary for the semantic content, but the representations themselves are encoded information that can be detached and abstracted from the systems that provided the input and allow for potential output.

To make sense of this, consider when a child first acquires the concept CAR. It may be based on visual, tactile, and auditory input received at the same time. For example, a toy car could be placed in a child’s hand, which she then looks at while her mother points at it and says, “Car.” This combined information will provide content for the child’s mental representation CAR, yet the representation itself will not be bound to the actual toy car she handled or the sensorimotor information from that one event. Instead, it is encoded and abstracted in such a way that later, when she sees a cartoon version of a car or dreams of riding in a car, her representation of CAR will be utilized even though it is being employed in a way that is distinct from the bodily experience that helped provide the content for it. Thus, in a certain sense, representations are independent from the body and its interaction with the world. If the original car is lost the child still has the concept CAR. If she somehow loses the ability to hear, see, or touch cars she still has the concept CAR. Even if she loses all possibility of communication she still has the concept CAR (presumably). Perceptual input is essential to forming a concept about an external object, but once it is encoded as a symbolic representation that information has meaning separable from the initial input and can be applied to other symbols in purely abstract, formal ways to generate more mental content not tied to perceptual input or physical output. This is similar to looking up a word in a dictionary. The meaning of a given word can be ascertained simply by how it relates to other words. Where the letters or words came from is of little importance and what, if any, objects the words correspond to beyond the page may be of little or no concern as far as meaning goes. This is partly why Robert Cummins remarked, "Cognitive science is founded on the empirical assumption that cognition (hence the study of cognitive systems) is a natural and relatively autonomous domain of inquiry." (Cummins, 1989: 19) The systems and symbols necessary for cognition are physical entities clearly affected by and affecting the natural world around us, yet a representational theory of cognition translates embodied physical interactions into a completely new, autonomous set of neural physical interactions in which representations are distinguished by their unique, computational semantics and syntax.

This autonomous, "classical sandwich" allows for brain-in-a-vat scenarios to be legitimate possibilities. Though many mental symbols and representations have content tied directly to external input, that input is only valuable to cognition once it has been translated into Mentalese and utilized via the language of thought. And, once it has been translated in such a way, it is not dependent on anything other than the syntax and semantics of the representations within the brain. Thus, there are at least two ways to imagine a brain-in-a-vat scenario where the brain has cognition. For the first scenario, a human being could live a relatively normal life, with a fully functioning body, but then have his/her brain removed for some reason (due to a traumatic accident, for example). So long as the brain were to receive the electrical stimuli, blood, oxygen, and other nutrients it needs to function, there's no reason to suppose the brain wouldn't remain cognitively active. The person/brain may not have the capacity for output in terms of oral communication or movement anymore, but thinking could still occur because symbol manipulation could still take place. For the second scenario, we have to imagine certain advances in neuroscience and technology. Supposing the semantic content of mental symbols is the result of very specific forms of stimulation to neurons or clusters of neurons, the input provided by specific sensory systems is based on how each system provides a unique sort of stimulation. Thus, visual representations operate as they do not because the visual system provides inimitable information, but rather because the visual system excites the brain in a certain way. If that excitation or stimulation could be duplicated via electrical stimulation of the necessary neurons in similar ways, then it should be possible to replicate visual representations without a visual sensory system. Though we aren't there yet, if neuroscientists could exactly duplicate the neural stimulation experienced via embodied sensory systems by using only targeted and controlled electrical signals (assuming electrical stimuli is the key) then a brain-in-a-vat, even if it had never been in an embodied form, could presumably have cognitive activity like that of an embodied individual. The fact that it would be false information does not matter. Cognition would still be taking place and all that matters is the brain, the symbols within it, as well as their semantic content and syntactic capabilities.

Additionally, identical neural activity to an embodied brain may not be necessary for a disembodied brain to be cognitively active. To understand why, we can consider Hilary Putnam's famous Twin Earth case. Imagine two embodied minds: one is Art's, on Earth; and one is Bart's, on Twin Earth. In Putnam's scenario Art has the concept WATER that applies to the clear liquid comprised of H_2O that fills the oceans and falls from the sky. Bart also has the concept WATER that applies to a clear liquid that fills the oceans and falls from the sky of Twin

Earth, but his "water" is actually composed of elements XYZ. Though they would both use the term "water," the term is used to express two different concepts because one concept corresponds to H_2O and the other to XYZ. Now, if we imagine a brain-in-a-vat stimulated to have a mental content similar to Art, it will be unclear whether the envatted brain will experience a concept closer to Art's or Bart's because the only way to determine whether the concept corresponds to H_2O or XYZ would require interaction with the external world. And since the envatted brain, or anyone studying it, will not be capable of establishing the truth of whatever WATER concept it possesses (if any) it cannot be said definitively that the envatted brain will have a mental life exactly like that of an embodied mind. This dilemma has lead to a distinction between wide and narrow mental content within the traditional accounts. Narrow content is the content that determines the psychological role of the concept WATER, and in the narrow sense Art and Bart (as well as an envatted and embodied brain) could be said to have the same mental content because the mental representation constituting WATER will have the same causal powers by playing the same semantic and syntactic roles in both minds. Presumably, this could be the same in any brain based solely on the types of symbols and the causal powers of those symbols within a given cognitive system. The wide content is where they would differ. The wide content is what comprises the relation of the narrow content to its real-world referent. So Art's wide content includes H_2O and Bart's wide content includes XYZ—mental content that could not be replicated without some embedded or embodied experience. This distinction of content is important to consider because wide content could play a significant role for embodied theories.

For the embodiment theorist, if wide content is deemed essential to cognition, it rules out the possibility of any brain-in-a-vat scenario, and refutes a large component of the mental sandwich depiction, because wide content is unattainable without worldly interaction. However, wide content doesn't necessitate embodied cognition, nor preclude disembodied cognition. It could be that a brain-in-a-vat lacks an essential feature of cognition because it lacks the ability to relate narrow content to objects outside the brain, but brains that have wide content due to having perceptual capacities to interact with the world might only have wide content due to causal relations with perceptual systems. That may sound like an obvious statement. But if the relationship between wide content and any mental content is only *causally* linked to embodied perceptual systems, then the traditional model can still account for it. The key for embodied theorists is to show that some wide content—and the perceptual systems necessary for it—*constitute* semantic and syntactic properties of mental representations. Otherwise, both wide and narrow content, operate according to the traditional model. Wide and narrow content could be considered as operating similarly in terms of the neural processes underlying them—wide just has the additional feature of mental content constituted by perceptual input. (And, as will be discussed in later sections, metaphysical constitution is a much more difficult claim to establish than nomological causation.) Of course, this is assuming wide content could be essential for cognition. If it's not, and instead plays a useful (but nonessential) role in providing empirical truth conditions for narrow content, then traditional cognitivists could point out that brain-in-a-vat scenarios do not need to produce identical mental content in order to be considered legitimate scenarios and the mental sandwich is still in place, just with one less layer in the middle.[2]

Another lingering point of ambiguity within this description of the traditional model is the "manipulation" of mental symbols. At minimum, manipulation has to do with the movement,

---

[2] Fodor initially supported the separation of narrow and wide mental content (1974), but later questioned the division as superfluous because "computational-syntactic processes can implement broad-intentional ones because the world, and all other worlds that are nomologically nearby, arranges things so that the syntactic structure of a mode of presentation reliably carries information about its causal history." (Fodor, 1994: 54)

handling, or operation of a part or parts. Sometimes this is due to the will of an individual, such as the manipulation of chess pieces during a chess match. And sometimes this is due to the necessary function of a part within a system, such as the hammer striking a string in a piano when a key is pushed. The former makes sense when describing the use of a language like English because one can pick and choose the words and sentences one needs to properly express a thought, and manipulation becomes a kind of means-end reasoning. However, when discussing thought itself, the manipulation of symbols and representations is not due to the will or desired ends of an individual because the will or desired ends are thoughts themselves. And since the traditional model does not want to postulate a sort of homunculus dictating the rules in which Mentalese is to be used, the manipulation must be something else. It might be more like the piano example. Not in the sense that someone is pressing the keys, but in the sense that the piano has a set of rules built into it. The piano is a system that, once constructed, has very specific ways in which it can operate, and predefined ways in which different parts of the system can work together (pedals, dampers, and strings, for example). Though this is mechanical, one can say there is a sort of algorithm governing the operation of the piano. Step on the pedal before pressing a key and a specific chord is played. Step on the pedal after pressing a key and that chord is altered. This is analogous to cognition in that the syntax of the language of thought is believed to be a set of limitations and possibilities hardwired, so to speak, in our neural structure that dictates the manner in which mental symbols can interact. Just as the molecular structure of the various parts of the piano, and how each of those parts is aligned with other parts, determines the possible interplay and sounds that result from those interactions, the chemical, electrical, and structural makeup of the brain defines the manner in which signals, symbols, neurons, and various neurological systems can properly work together. This innate syntax can be considered an incredibly complex set of algorithms that, starting from the initial neurochemical state(s), determine what input is allowed, the form the input must take in order to be processed, the many potential paths the process could take, as well as the output resulting from those processes. This is what allows for both reducing and compounding concepts (such as having the concept DOG and reducing it to constituent components of MAMMAL, HAIRY, etc., or combining the concepts HORN and HORSE to create UNICORN) and for all forms of inference.

Thus, the reason the little girl in the example with the car is able to associate the toy in front of her with her concept CAR is because her cognitive system includes algorithmic-like rules that make a connection between the symbolic representation she already possesses for CAR and the symbolically encoded information received via her retina. She may be consciously aware of some of this in that she may think propositionally about the traits a car should have and the traits the toy in front of her has, and if the two should fall under the same category. But such propositional attitudes and exercises of reason are the result of the underlying syntax, not the sole expression of it.

It is also very important to note here that the syntax operates sequentially in most traditional models. Just as with the piano or the way words in a sentence must be read in a certain order to function correctly, cognition has a definite order of events. It may be remarkably fast—on the scale of nanoseconds—but processing within any cognitive or perceptual module is not instantaneous. A specific process occurs and must be followed in order for cognition to happen. Just as the thought expressed in this sentence loses meaning if the words are read right to left rather than left to right, with subjects, objects, verbs, prepositions and such in specific positions relative to one another, Mentalese functions successfully only when mental symbols are organized correctly and processed in a certain order.

Now let us briefly consider how this relates to the specific mental process of visual perception. Firstly, the syntax mentioned above is responsible for one's ability to make inferences in regards to perception. The input received via the perceptual systems is impoverished and does not provide objective information about the world to the brain. Retinal images, for example, are underdetermined. When looking at an object, it is perceived from a vantage point that may not reveal its objective shape. At certain angles a penny can look like an oval rather than a circle, a square can appear not to have four 90-degree angles, a can may not look like a perfect cylinder, etc. And yet, the penny is determined to be a circle, the square to have four 90-degree angles, the can a cylinder, and so on. The perceptual input alone is not enough to reach these conclusions. Instead, it is believed that the actual shapes of the objects are derived from retinal information combined with assumptions about the world based on mental concepts and rules governing the possibilities of those concepts. Thus, assuming one already possesses the concept of PENNY, it likely includes semantic content regarding size, color, shape, use, and other distinguishing features of what makes a penny a penny. If it is believed that all pennies must be circular, then any object that is square would automatically be excluded from consideration as a penny. However, even if the penny is observed from an angle and appears to be oval, one may also have concepts regarding changes in shape due to changes in perspective. The rational mind, with its language of thought, infers a connection between these concepts (as well as concepts and propositions regarding the body, movement, the environment, and other things) that allows one to reach the conclusion that the object may in fact be circular when perceived from another perspective. One way of looking at this is that, since this is all brain-based, such a conclusion does not require one to physically investigate the penny to learn if the hypothesis is correct. The cognitive feats necessary to reach the conclusion that might spur bodily engagement with the world and a definitive conclusion about the shape of the object—though about something in the external world—occurs entirely within the mental realm. So, once again, even if the process relies on certain information related to objects in the world, the cognitive processes utilizing concepts and inferences, and the innate syntax that govern both, occur separate from the sensory systems providing the input and the motor systems that may demonstrate the output.

This may sound as if traditional accounts are not concerned with sensory systems. This is far from the truth. The last example with the penny involved a case in which one already has the concept PENNY, but there must still be an explanation for how one comes to develop such a concept from visual experience, given that visual information is underdetermined from the outset. David Marr provided such an explanation (Marr, 1982). For Marr, vision is basically a complex, representational information-processing task from which one tries to reliably derive properties of the world. The most important aspect of Marr's model was his explanation of different levels of processing that occur in vision. Rather than assuming that semantic content of concepts alone helps one to evaluate underdetermined visual input (as presented above), Marr says the ability to discern ovals from circles and recognize ovals as circles at different angles, is built directly into the visual system itself. This is due to three levels of representational processing:

1. The *primal sketch*, which is mainly concerned with the description of the intensity changes in the image and their local geometry, on the grounds that intensity variations are likely to correspond to physical realities like object boundaries.

2. The *2½-D sketch*, which is a viewer-centered description of orientation, contour and depth, and other properties of visual surfaces.

> 3. The *3-D model*, which is an object-centered representation of three-dimensional objects, with the goal of allowing both handling and recognition of objects. (Poggio, 1981: 258)

Only at the 3-D model level does one begin to store new descriptions of shapes that can then be categorized with or deemed contrary to previously stored representations of three-dimensional models. This categorization is possible due to hierarchical processes and descriptions within the 3-D model level. Like an informational pyramid, the lowest levels have the most bits of description. In regards to the penny, the lowest level could include the many subtle shapes, angles, shadows, etc., found on the penny due to the words and shapes imprinted on it—the things that would separate a penny from a dime, for example. As the levels go higher, the information carried in the descriptions become less and less. This could go up from color, to size, to basic shape—all of which contributes to how one categorizes the perceived object. And all of this happens in the process of perception, so one is not necessarily aware of this categorization. Thus, one might approach a penny lying on the ground and realize that, though it appears as an oval at a certain angle, it is actually a circle because the perception shares information with other perceptions of circles—and this occurs without ever inferring the circularity of the penny from the semantic content of the concept PENNY. Rather than categorizing from the concept outward, by comparing new input to existing concepts, the visual system automatically differentiates input. In this way, Marr's computational process of perception is imbued with discriminatory abilities that allow for the forms of categorization that are essential for one's cognitive capacity to separate objects and develop unique concepts for everything that can be seen in the world.

Initially this may sound as if it's connecting a computational model of cognition to the world in a way that goes against the classical sandwich, but not really. The relationship between the world, perceptual systems, and cognition is still only causal. If combined with Fodor's language of thought, Marr's perceptual model provides an account of how the subtle variations in semantic content and syntactical ability come to exist in mental symbols based on visual input. The symbols, what they mean, and how they can be used is *caused* by the perceptual system, but not *constituted* by it. Once discriminated and categorized, the 3-D model representation of a penny does not continue to rely on the visual system in order to be used in cognition. Though the concept type PENNY could not exist without embodied interaction with the world, later tokens of it can. And in such instances, those tokens are now independent of the visual system even though everything defining that mental symbol as a PENNY finds its origin in visual perceptual system.

There is much more that could be said about the traditional approach as a whole. However, this should suffice for the purposes of this work. The obvious issue(s) at stake in this debate revolve around the "mental sandwich" and the causal/constitution relation of the parts of the sandwich. For example, according to Fodor, a model involving an innate syntax which governs the manipulation of mental symbols and representations at least means "thought is prior to perception (because perception is, *inter alia*, a kind of inference)…concepts are prior to percepts (because inference requires, *inter alia*, subsuming a percept under a concept)…thought is prior to action (because acting requires planning, and planning is a species of reasoning)…[and] action is the externalization of thought." (Fodor, 2008:12). Each of these directly or indirectly supposes a model of cognition *causally dependent* on embodied interaction with the world—otherwise there would be no percepts or action or semantic content, possibly—but does not present a model in which embodied interaction *constitutes* cognition to the point that cognition as a whole cannot possibly exist without it. This distinction will continue to be at the heart of much of what is to follow. As we will see, many embodiment theories fail to refute the possibility of conceptual types existing independent of conceptual tokens, which will allow for a certain type of disembodied cognition. But not all. Some theorists are making radical claims that would require a complete rethinking or abandonment of the core of traditional cognitive science.

3. The Case for Embodiment

3.1 Mind, Body, and World are "Coupled."

The most common central theme to all theories of embodied cognition is that mind, body, and world are uniquely linked. Rather than being three distinct things, they exist in such a way that one cannot be understood without its relation to the other two. This is often captured with the term "coupled." Many analogies and examples may come to mind with this term: train cars have couplings that allow their movement to be linked, if two people are committed to one another romantically they are said to be a couple, and software engineers may couple computer programs so that the module of one program relies on the module of another program. In each of these examples, when coupling occurs a new system is created by the joining of two things that could have been, or were, separate before. This new system will require a different set of parameters to explain it compared to when the coupled objects were independent from one another.

Consider the motion of a lone train car. Say you want to know what force would be needed to move the car one mile in 30 minutes. Even if you were unsure of the track ahead—whether it was flat, or on an incline or decline—you could calculate the force needed and provide answers for various track scenarios just by knowing the specifics of the train car (plus a little knowledge of physics and gravitational force). Other freight cars that are not linked to the freight car in question do not factor into the equation. However, when it is coupled with other freight cars and locomotives this is not the case. When coupled, the force needed to move that specific car can no longer be tabulated by considering the car on its own. The weight of all the other cars must be considered; the power of the locomotives' engines become a factor; and, since the total line of coupled cars could span across flat, inclined, and declined tracks, the current position and path of all the cars must be included in the equation. If the majority of the cars are on an incline, more force will be needed. If many of the cars are on a decline, less force will be needed. Obviously, this can still be figured out—as it is done daily in the shipping industry—but the effort to do so is not simply a combination of a bunch of isolated equations regarding each train car. The ability to answer how much force is needed to move the one car when it is coupled with other train cars and locomotives requires one to consider the train as a complex, interlocked, interdependent system of cars and engines where the motion, movement, and force of one has an effect on the motion, movement, and force of all.

However, there is a flaw in this coupling analogy when relating it to embodied cognition. Train cars can be decoupled and exist independently, whereas most embodiment theories seem to say that if mind, body, and world are decoupled they cannot exist independently. Sure, a brain (considered as a mass of grey matter) can exist without a body, but it will not function as it did when coupled—it will just be a bunch of cells grouped together, but it won't have a mind. Similarly, a body without a mind is really just a mass of insentient cells. Generally, when "body" is referenced in discussions of mind it is an object that is both housing and directed by consciousness. Thus, if the mind is removed it's no longer a body in this sense; it is just a collection of various tissues and cells. To have clearer idea of what is intended and entailed by "coupling" in theories of embodied cognition it's best to look at the earlier uses of the term as applied to cognition, and how some contemporary philosophers are using or have used the term.

Though Maurice Merleau-Ponty did not use the term "coupling" explicitly, he is undoubtedly one of the earliest and most influential proponents of mind-body-world interrelatedness in cognition. As a self-proclaimed phenomenologist, he attempted to examine cognition from a first-person perspective in order to challenge the more empirical, objective approaches that had preceded him and were prevalent during his lifetime. In *Phenomenology of Perception* (1962), he set forth a number of bold claims and arguments in an effort to describe the nature of one's perceptual contact with the world. Merleau-Ponty was against both traditional empiricism and rationalism because he thought neither could effectively describe the experience of being a perceiving agent in the world. Both approaches suppose a sort of dualism between consciousness and the world wherein the body is treated as just an object controlled by the mind that allows an individual to interact with and extract information from the outside world. In the case of dualism, the mind is independent from the body and world. In the case of empiricism, the world is independent from the mind and body. When looking at the experience of perception this is clearly not the case, according to Merleau-Ponty. Rather than the mind affecting the world via the body and vice versa, there is a sort of continuous reciprocal causation[3] wherein the body is not a mere conduit of external input and internal output. Instead, the body-mind—as a coupled system—is both perceiving and bringing about the world in which it is perceiving; the body-mind perceives by responding to changes in the world (in duration, motion, light, etc.) while the world one is perceiving is changing due to one's embodied engagement with it. Merleau-Ponty used an example involving trying to hold a hamster with tongs to make this point:

> When my hand follows each effort of a struggling animal while holding an instrument for capturing it, it is clear that each of my movements responds to an external stimulation; but it is also clear that these stimulations could not be received without the movements by which I expose my receptors to their

[3] This phrase, "continuous reciprocal causation," is used by Andy Clark to describe how one's actions are continuously responsive to events in the world that are, at the same time, continuously responsive to one's actions. However, he specifically notes that he believes Merleau-Ponty stressed the same idea. (Clark, 1997)

> influence…The properties of the object and the intentions of the subject are not only intermingled; they also constitute a new whole. (Merleau-Ponty, 1942: 13)

The intermingling of which he is speaking is in line to the more recent use of "coupling." In the example with the hamster, his hand moves in response to the movements of the hamster and continuously responds to how the hamster's movements respond to his own. The hand—or body—in such a scenario is not solely a medium through which the mind is receiving input and then dictating output; it is not just an object feeding information from the world to the mind and from the mind to the world. Rather, it is an integral part to the relationship between the two. The perceptions that one is receiving while trying to capture the hamster are not processed and reflected-upon to generate conscious, propositional ideas on how one could and should move one's hand in order to properly control and contain the hamster. Instead, the perceptions are tied directly to sensorimotor input and feedback, as well as the cognitive intention of capturing the hamster. Thus, when one's locomotion is based on some cognitive goal, such as keeping the hamster still, "we notice for the first time, with regard to our own body, what is true of all perceived things: that the perception of space and the perception of the thing, the spatiality of the thing and its being as a thing are not two distinct problems." (Merleau-Ponty, 1962: 171) They are not distinct problems because they are coupled.

However, Merleau-Ponty's approach and example are focused on the phenomenality of coupling. That is, when examining the experience itself there is no discernible distinction between the motion of the hand and the goals of the mind. It seems to be an instantaneous response to the movements of the hamster. In such experiences the body isn't separate from the cognitive goals dictating its action. Body and mind are dynamically engaged with the world to the extent that one cannot find the point where cognition begins and body ends, and vice versa.

Of course, a phenomenological description of coupling alone would likely leave most cognitive scientists unsatisfied since it is inherently limited to subjective experience and does not strive to provide quantitative, objective evidence of coupling. One of the more notable recent uses, which at least aims to provide a much more objective description than Merleau-Ponty, is in dynamical systems theory.

In a very broad sense, dynamical systems theory is just attempting to describe how a system changes over time. In terms of a cognitive system, this means being able to explain how an infant just learning to track moving objects can develop into a child who can understand natural languages, then to an adult who can discuss abstract theories of thought itself. Contrary to the traditional model, dynamical models argue that cognition and cognitive development are not organized by an innate neural syntax, or language of thought. They deny the existence of solely brain-based rules that determine how sensorimotor input is processed and stored so that it provides the conscious agent with a warehouse of information continually built upon and later utilized when needed. Instead, they say cognition is based on an agent's dynamic interaction with the external environment—interaction that can be achieved only through embodied activity. "[Cognitive] development can only be understood as the multiple, mutual, and continuous interaction of all the levels of the developing system, from the molecular to the cultural." (Thelen & Smith, 2006: 258). In terms of modeling and analysis, this means "systems, or parts of systems, are *coupled* when the mathematical description of the behavior of one must include a term that describes the behavior of the other." (Shapiro, 2011: 213; my emphasis)

To illustrate this last claim in physical terms, Timothy Van Gelder has used Watt's centrifugal governor to explain how a system can function dynamically coupled rather than through a clearly delineated series of rule-governed, causal events (1995). Shown here, the governor does not function in a step-by-step, algorithmic manner where input from one part of the system determines the actions of another part of the system. Instead, the governor is constructed so that steam pressure through the valve pushes the flywheel, which extends the spindle arms at the same time,

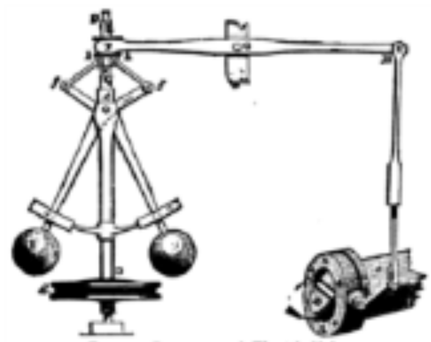

Figure 1 Watt's Centrifugal Governor

which adjusts the valve at the same time, which pushes the flywheel at the same time, which causes the spindle arms…and so on. Changes in the system are interconnected and instantaneous. The idea being that there is no order of operations or processing and no clear starting point to the causal process other than stimuli from outside the system (in this case steam).

Though this example is clearly far less complex than the human mind, the hope is that it illustrates how a system, such as cognition, could be coupled to the body's engagement with the world, implying that any attempt to describe cognition must also factor in the behavior of the body. The reason why this can be considered a coupled system is because "factoring in" the behavior of the body doesn't operate in a causally linear fashion where one simply considers what is occurring to the body in order to account for the mental state, or vice versa. Instead, the factors, or components of the system continuously interact with each other in a nonlinear fashion, affecting each other and thus the system as a whole simultaneously. "The coherence [of the cognitive system] is generated solely in the relationships between the organic components and opportunities of the environment." (Thelen & Smith, 2006: 281). In the traditional model, coherence is generated from the syntax built into the neurological system. Fodor claims thought is prior to perception. Interaction with the external world causally affects cognition, providing semantic content it could not have otherwise, but the language of thought takes causal priority in terms of the organization of that input or the generation of motor output. The dynamical model proposes *multicausality*, wherein “no single element—internal or external—has causal priority.” (Thelen & Smith, 2006: 281)

In addition to van Gelder, Esther Thelen and Linda Smith also used dynamical systems theory to describe the behavior of infants in the A-not-B task (Smith & Thelen, 1994), and Randall Beer believes a computer experiment involving an artificial agent’s ability to detect the difference between circles and diamonds provides evidence that our nervous systems (including the brain) are necessarily embodied systems dynamically engaged with the environment in which we are situated (Beer, 2003). More of this will be covered in later sections, but the key takeaway for now is that each of these theories holds that descriptions of cognition—whether focused on psychology, like Thelen and Smith, or artificial intelligence, like Beer—cannot be successful without a description of the agent’s interaction with its environment. In contrast with Merleau-Ponty, these descriptions of interaction are not based on first-person accounts. With someone like Beer, the interaction is a describable and/or quantifiable factor of the system. Though this may not entail the rule-governing syntax that forms the backbone of computational approaches, the belief is that interaction between and/or within the system(s) over time can be visualized and analyzed in an objective, scientific manner. This is achieved by looking at “coordinated patterns of behavior, rather than the sequential sense-think-act processing style that is typical of computational approaches.” (Beer, 200: 97) The manner of calculation needed to establish the patterns of dynamic systems and the states which constitute them are exceptionally complex—involving differential equations, vector fields, limit sets, and other factors—so I will not attempt to explain precisely how the equations operate. However, the conclusion from Beer clearly echoes the sentiments of Merleau-Ponty.

In the excerpt that follows, Beer is describing how a simple artificial agent was programmed to correctly categorize falling diamonds and falling circles, while attempting to catch the latter and avoid the former:

> As the agent's state evolves from its current point, the agent's resulting actions and the environment's own dynamical evolution change the sensory input that the agent receives and modify the subsequent trajectories that are available to it. In this way, both the agent's dynamics and that of its environment continually shape the unfolding behavioral trajectory, as well as the future sensitivity of that trajectory to subsequent sensory input. (Beer, 2003: 236)

Just as Merleau-Ponty had said "the spatiality of the thing and its being as a thing are not two distinct problems," the internal goal of the agent—to catch and avoid certain shapes—that is providing its being, so to speak, is not distinct from the external events. What the agent will do (its behavioral trajectory) is coupled with its particular form of embodiment and situatedness because it is continuously affected and shaped by what the agent perceives from its environment.[4]

Another famous and influential account of coupling comes from Varela, Thompson, and Rosch (VTR from this point forward). Their description of coupling is revealed in their paradigm example about the experience of color. They believe color to be uniquely suited to explain coupling and embodiment because it provides a microcosm of cognitive science—as our understanding of color has been informed by neuroscience, psychology, artificial intelligence, linguistics, and philosophy—and because the experience of color has a significant effect on

[4] Abstractly, the issue of continuous interaction/affectedness within the system is not necessarily tied to external variables because the requirements of a dynamical model do not entail one that must causally interact with the external world. However, when discussing human cognition or attempts at artificial intelligences intended to replicate human-like cognitive abilities, the external world becomes a necessary variable, as there is the assumption that cognition involves at least the variables of contained organic systems and subsystems situated in a broader environment.

human experience. The main traditional claim they take issue with is that colors exist in a pregiven world, independent of a perceiver. If that were true, and each color were understood, say, as identifying some specific wavelength of light, then not only would every color have an objective standard against which it could be judged, it would also imply that experiences of certain colors would correspond to experiences of certain wavelengths of light. Assuming two persons' visual systems were functioning similarly and normally, when looking at the same STOP sign each person would be experiencing similar wavelengths of light and experiencing them as RED. And, even if someone's visual system was malfunctioning, and he or she were red colorblind, experiencing the light reflected off the STOP sign would still yield a consistent experience each time (even if it weren't red). The color of a given object is what it is because of the wavelength(s) of light reflected off the surface of the object. We receive and process this visual input and correlate it to color concepts. If the light input is different, then presumably so too is the color concept brought to mind. VTR believe this is wrong.

Unlike a computer system, in which any sort of coupling with the external environment is based on input/output relations, the coupling that occurs in an organism's experience of color is based on the organism's perceptual capacities and the unique history of activity it has with the external environment. Because of this, color can't be explained if located independent of our perceptual capacities. Instead of viewing perception and color as separate but interacting, "we must locate color in the perceived or experiential world that is brought forth from our history of structural coupling." (VTR, 1991: 165)

To explain the unique coupling that occurs in color experience, VTR look at both the phenomenology and physiology of color perception. Colors are usually described as combinations of the six basic colors of blue, red, green, yellow, black, and white with variations in hue, saturation, and brightness. One may see a light red, a deep blue, or a pale blue-red (purple). At first this may seem like an arbitrary categorization in an effort to breakdown the various colors in the world. However, it may have a more psychophysical source. In 1957, Leo Hurvich and Dorothea Jameson published the *opponent-process theory*. This theory claims the visual system is composed of three color channels, one of which signals differences in brightness and two of which signal differences in hue. These channels are believed to correspond to complex cross-connections among retinal cells and postretinal neuronal ensembles. The importance of this theory, for the purposes of this example, is that it asserts the three mosaics of cone cells in the retina—long-wave (L), middle-wave (M), and short-wave (S) receptors—process colors based on opposing interactions between them.

> The difference between the signals from the L and M receptors generates the red-green channel, and the difference between the sum of the signals from the L and M receptors and the signals from the S receptors generates the yellow-blue channel. These two chromatic channels are opponent: an increase in red is always gained at the expense of green and vice-versa; an increase in yellow is always gained at the expense of blue and vice-versa. (VTR, 1991: 159)

This idea becomes particularly interesting when determining the color of a given object in isolation versus when other objects surround it. One might assume the color would be perceived the same, but that doesn't seem to be the case. The checkerboard illusion by Edward Adelson (1995) is an excellent example of this phenomenon. Square (A) appears to be darker than square (B). However, if rectangles of the same color are added connecting the two, one can see that the squares are actually the same shade of gray:

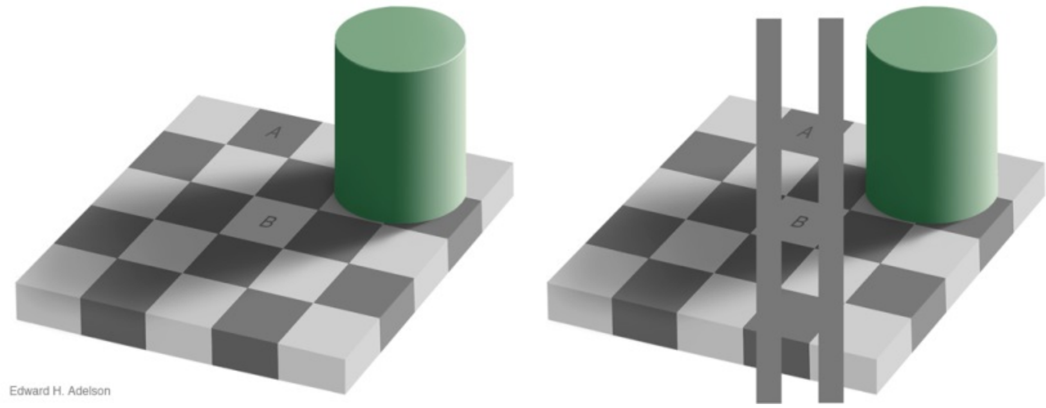

Figure 2 Adelson’s checkerboard illusion

This is not due to a flaw in perceptual ability nor is it unique to the color gray. (If you were to swap the gray for shades of green found in the cylinder in the image, the same misperception would occur.) In the Adelson example, the color is the same but we believe it to be different due to tricks the visual system plays when attempting to differentiate the shaded from non-shaded areas. Additionally, there are instances where one perceives a color to be the same when an object is isolated (as with the right-hand checkerboard above) as well as when it is surrounded by other objects. Yet the light perceived is actually different.

Take something green, like a pepper. When the pepper appears green it typically means it reflects a high percentage of M light and a lower percentage S and L light. Thus, it would be reasonable to conclude that one’s eyes are receiving more M light reflected off a green object when looking at it because the M light is what allows one to perceive it as green. But the perception of M light may not be sufficient for this perception. If we were to take the pepper and put it in a more complex scenario, surrounded by many other vegetables and fruits, the pepper will actually reflect more S and L light (short- and long-wave light) than M light. But that doesn’t stop one from perceiving and categorizing the pepper as green in color. Even when surrounded by a red tomato, a yellow banana, an orange carrot, and a purple onion, the pepper is still clearly green even though the light reaching one's eyes is now S and L light, as compared the higher percentage of M light reflected when the pepper is on its own. If a high ratio of M light were essential for the color green to be perceived then it seems bizarre that one should still be able to see the pepper as green when it is reflecting a higher ratio of S and L light, as opposed to M. Therefore, VTR say that when the pepper is not on its own the light reflected locally from it is not sufficient to predict the perceived color of the pepper. And if the locally reflected light is not enough to guarantee a certain color perception, this somewhat explains the misperception in the Adelson example. The light reflected is the same and yet the perceived colors are different because locally reflected light alone cannot determine the perceived colors of the squares. As VTR say, "There is no one-to-one correspondence between perceived color and locally reflected light." (VTR, 1991: 160)

But if color experience does not correspond identically to the reflected light, what is it that provides consistency in both group and individual color experience? This is where the history of the organism and its particular perceptual systems come into play. Part of the reason why an object can be deemed green even when the light waves reflecting off it may change is because there are more systems participating in color perception than just retinal receptors. In primates, for example, subensembles of neurons in the thalamus, primary and extrastriate visual cortex, inferotemporal cortex, and frontal lobes have been demonstrated in color perception.[5] It's believed these subensembles are an elaborate patchwork of visual modalities that together constitute visual perception. Even though they have a degree of independence from one another, these different modalities—which may account for form, surface property, three-dimensional spatial relationships, and three-dimensional movement—"are emergent properties of concurrent

[5] In VTR, 1991: 161-162, from Gouras, P., and E. Zenner, 1981. Color vision: A review from neurophysiological perspective. *Progress in Sensory Physiology* 1: 139-179.

subnetworks, which…cross-correlate and work together so that at almost every moment a visual percept is coherent." (VTR: 162)

Color perception may just be one of these modalities, and it may function based on indicators of hue, saturation, and brightness, as mentioned before, but VTR claim it can only be understood—or create color perception and understanding—depending on its relation to the other visual modalities. Though the mutual effects of the relationship the retinal receptors have with the rest of the visual system is still speculative, VTR believe it may be able to account for the perception of green even when the surface of the pepper is reflecting less M light and more S and L light. Since the receptors are interrelated with more perceptual and neuronal systems, there is no "pure vision" that receives the light input directly, free from the influence of other perceptual modalities. Instead, the visual modality that includes the receptors is affected by—or at least processed simultaneously with—other modalities that may include background knowledge regarding the expected color of peppers, inferences about the light being reflected (e.g., expecting it to be darker if in shade and lighter if in direct light), and possible sensorimotor expectations regarding how the light/color should change based on one's movement. In this way, multiple modalities are believed to work together and assign colors to objects. Thus, color experience is not a matter of taking in information purely from a pre-given world through a single visual system. "Perception and action, sensorium and motorium, are linked together as successively emergent and mutually selecting patterns…Our colored world is brought forth by complex processes of structural coupling." (VTR: 163-164)

As a subtle point of clarification, VTR are not claiming the physiological visual system made up of the three types of cones is coupled, per se. That specific system can be damaged and the organism or agent can still perceive things other than colors. Likewise, it could be functioning normally, yet the agent could have a neurological disorder that does not allow the input from the visual receptors to be processed normally (as is the case with something like face blindness). But in the sense that it operates so as to allow the organism to consciously receive the visual input and make sense of it—that he/she/it can actually *perceive* colors—it is coupled in the manner described. Color may be just one form of perception, but VTR believe further investigation into all other forms of perception will yield similar results—such as auditory input categorization tied to background knowledge of certain sounds, or gustatory experiences tied to smells, sights, and expectations based in background knowledge (as well as the input from taste buds). Oliver Sachs and Robert Wasserman provided one case VTR reference in support of this last example. In their case study a patient had acquired cerebral achromatopsia and had become completely colorblind. This total loss of color perception, VTR believe, showed evidence of coupling by the effects this loss had on other perceptual systems and mental states. Aside from generally describing the world as appearing "dirty" due to its constant black and white appearance, the patient also "found foods disgusting and sexual intercourse impossible." Presumably, this was due to the complex structural coupling underlying the patient's perception of the world and "when these processes are altered, some forms of behavior are no longer possible." (VTR: 164) If categorization of a given object is dependent upon that object's appearance, and it's appearance is coupled with visual perceptions and sensorimotor possibilities (that are coupled together as well, with one another and background knowledge), then altering the appearance in a significant way—such as losing its color—results in the object being experienced differently. A key aspect of this example is that no objective traits of the object have been changed. If, say, the patient previously loved bananas but was now disgusted by them, the nutritional content, mass, size, and density of bananas remain basically the same, but the experience has changed due to traits that are coupled with other sensory modalities. In this case it was color, but a similar result could have come from changes in smell or texture as well. Though the objective traits just mentioned could be considered aspects of a pre-given world, those traits are necessary but not sufficient for developing concepts essential to cognition or describing the phenomenological experience of the banana. The world as seen and experienced is categorized and conceptualized in ways which bind information from different sensory modalities together is such a way that uncoupling them fundamentally alters the concepts gleaned from that experience.

If true, this means that concepts—regarding color determination, food and its appeal or repulsion, and presumably every other experience or potential experience—is constituted by this coupled information rather than just being the end result of a causal chain that builds upon the information from these different modalities. No matter which route we explore, VTR believe we will find that mind, body, and world are coupled because each is fundamentally shaped by the other(s) and none can exist independently.

By no means are these the only philosophers and neuroscientists using the term "coupled" in reference to embodied cognition theories, but these represent the cornerstones for all who are. They, unfortunately, are not nicely coupled together so that one person's use of the term is necessarily related to another's. It is possible, and quite common, to find one of these versions of coupling utilized in one theory but not others. For example, one need not be concerned with the introspective account provided by Merleau-Ponty in order to support the quantitative version supported by Beer. Yet in each case there is still the common thread that the system or experience being described and analyzed cannot operate at all, or at least not as intended, if factors of mind (or programming), physical form, and world are not intertwined. And any attempt to consider mind, body, or world in isolation will fail, as each one seems to be describable and possibly constituted by its relation to the others. In the sections that follow we will see what this means for specific aspects of cognition, from perception to language to mental representations.

3.2 Perception is essentially perceptually-guided action

There are certain things that seem to be a cost of entry in terms of existence for human beings. One is occupying space, another is at least the potential to move about in the world[6], and a third is the ability to perceive. The first two are fairly straightforward, but the last can be exceptionally difficult to define. At minimum, perception is understood as the interaction between sensory organs and external environment. For example, in grade school students are generally taught about the five primary modes in which the body perceives: sight, hearing, taste, touch, and smell. Via each of these modes of perception the body takes input from the outside world and sends it to the brain. The brain processes and deciphers the input and is able to populate the mind with information about the objects in the world. According to traditional cognitive science, this, combined with information previously acquired via similar means, allows the perceiver to make inferences about the input in terms of what one is actually seeing, smelling, touching, or so on, and reach conclusions about the external world.

In a very rudimentary way, when I smell something burning I initially take input through my olfactory receptors, which is then sent to my brain as smell x. My brain is then able to connect smell x with a previous experience of smelling something that had in fact burned, and I am then able to infer that the smell perceived might be something burning. I can then choose to investigate the smell further to find the source or ignore it and hope that nothing is set ablaze. This is an oversimplification, but the key thing to keep in mind, for now, is that in this presumed process (involving the mental sandwich as Hurley described) I am basically passive to the perception. The smell comes to me without my having to choose to engage with the world; the

---

[6] Even if one is paralyzed, the very notion of "paralysis" as a condition reveals the norm of locomotion.

world impacts me, and then I choose how to respond to it. If embodied cognition is correct, the idea of passive perception may be terminally flawed.

As mentioned before, taking up space (because of our bodies) and the ability to move about the world (with our bodies) seems to be a cost of entry for existence. Because of this seemingly unavoidable embodiment and subsequent interaction with the world, some have claimed perception is an inherently active process and experience. “Perception is not something that happens to us, or in us. It is something we do.” (Noë, 2004: 1) Accordingly, perception is portrayed by some supporters of embodiment as inextricably connected with the body in such a way that what we perceive is not objective information about the external world. Rather, it is information filtered and formed by the specific types of bodies we have and the potential actions of those bodies. Additionally, it may be information we can continue to possess only if we retain the sensorimotor systems responsible for this information.

This has many possible implications. It can mean inferences the mind is able to make are based on sensorimotor possibilities rather than formal associations between mental symbols and representations; it could also remove the need for cognitive inferences about bodily actions; and it may entail that reality itself is shaped and defined by our embodied interaction with the world. Regardless of where each theory ends up, nearly all philosophers confronting such issues in perception reference J.J. Gibson as the voice that began to change the discussion from cognitive computation to cognitive embodiment.

At the core of Gibson’s research and stance is a challenge to the notion of underdetermination and the reliance upon inference in perception. To make sense of this, consider retinal images. When looking at a given scene a perceiver registers many images via the retina. Right now, for example, I can see a computer monitor in front of me as well as the desk on which it sits, a wall behind that, a window to the right, and a stapler to the left with a pile of books behind it. I perceive these objects due partly to the light reflected from them and due partly to the shapes the reflected light seems to reveal. The monitor looks like a rectangle, the books look like trapezoids, and so on. Two issues with inference are occurring here. For one, most books aren't trapezoids; they're rectangles. My ability to associate the perceived trapezoid with the concept of a rectangular book is based (at minimum) on an inference relating to my experience with three-dimensional rectangles viewed from different angles and the background knowledge that books are often found on desks. That is, I have experienced seeing books from directly overhead, from the side at a 60-degree angle, from a 20-degree angle, and so on. In each of these instances I have associated each perceived shape with the concept of a rectangular book and awareness of my change in position. Due to this I have inferred that it has been my position that has changed, not the shape of the book(s). Thus, I do not have to amend my concept of "book" to include myriad possible shapes such as rectangles and trapezoids. Instead, I infer that rectangles can appear as trapezoids given certain changes in the position of perception. Additionally, I have often found books on desks. So that when I perceive a three-dimensional rectangle on a desk, if it is of a certain size, I may presume it to be a book rather than another object.

The second issue is that my perceptions alone underdetermine what objects are around me. Many objects other than books could present the same trapezoidal shape to my retina. It could be picture of a pile of books or a trompe l'oeil drawing on my desk. This is why many optical illusions are possible—at the right angle, a two-dimensional shape can look three dimensional, or one three-dimensional object cannot be distinguished from another. Such optical illusions can occur because we make inferences about the objects in front of us based on the limited information available to and from our visual receptors. The input from retinal images alone is not enough. Instead, as described in the overview on the traditional model, visual perception is able to function effectively by combining the retinal input with computational inferences and processes that allow a person to consider the book as a rectangle, and estimate the approximate size of it and the distance it is from one's current position. Gibson disagreed with this approach.

Where the traditional model holds visual information to be underdetermined, Gibson's ecological theory of perception denies this in favor of direct perception. Rather than the brain receiving input from the retina and then completing the picture, so to speak, removed from the visual system and external world, Gibson's theory takes vision as an embodied process wherein motion and interaction with the world provide visual information that is not underdetermined or impoverished. If we return back to my view of the computer, desk, books, and stapler, the information I am receiving while seated is what Gibson refers to as the ambient optic array. "To be an array means to have an arrangement, and to be ambient at a point means to surround a position in the environment that could be occupied by the observer." (Gibson: 65) An ambient optic array is the point at which light converges to provide an image of the environment—in this case, my retina. As I am seated, I believe the particular shapes presented to me to be books, a stapler, etc. If I stand, the light reflected from these objects reveals changes in the shapes of objects around me. As I get closer to looking straight down on it, the angles of the trapezoid-shaped book begin to more closely resemble those of a rectangle. The angles of the rectangular computer monitor become more trapezoidal as my perspective shifts from being directly in front of it to looking down at it. And, some objects that were previously hidden—such as an outlet on the wall just below the desk's surface—become visible. The outlet is taken to be a new object in the space around me; however, the books, monitor, and desk are taken to be the same even though their shapes have changed. According to the traditional model, we are able to infer the objects are the same and that it is merely our perspective of them that has changed—almost as if we take in the world through a series of snapshots and infer the connections between them. Gibson, on the other hand, believes perception incorporates time and space without the need for an inferential glue to hold it all together.

> The geometrical habit of separating space from time and imagining sets of frozen forms in space is very strong. One can think of each point of observation in the medium as stationary and distinct. To each such point there would correspond a unique optic array…This is an elegant and abstract way of thinking, modeled on projective geometry. But it does not allow for the complexities of optical change and does not do justice to the fact that the optic array flows in time instead of going from one structure to another. What we need for the formulation of ecological optics are not the traditional notions of space and time but the concepts of variance and invariance considered as reciprocal to one another. (Gibson: 74-75)

By variance and invariance, Gibson is referring to objects that remain constantly present while one's position changes. As I stand, the desk, books, stapler, and monitor all remain within my field of vision even though certain aspects of their appearance change. These are invariants. The outlet, however, varies. It comes and goes depending on whether I am sitting or standing. The relationship between variants and invariants is important because it provides information about the world in a manner that is not underdetermined at all. For example, since the desk only changes slightly as I stand, it is providing information about its approximate size and spatial relationship to me, and the fact that it remains constant while the outlet comes and goes provides information about the outlet's size and nearness to the desk (if it were farther away or much larger, it would not come into view by my perspective changing only about two feet vertically). In this way, Gibson believes that invariants in our visual field provide a great deal of information about variants and vice versa.

The horizon is a paradigm example of an invariant because it can provide information about one's position (such as being right side up or upside down) as well as one's nearness to another object, depending on how it moves with relation to the horizon. Similarly, when riding 20 miles-per-hour on a bicycle, a tree 300 meters away will vary much less than objects only a few meters away, which will vary often and quickly. And, if vision functions fluidly through time, rather than as a series of individual snapshots, the relationships between variants and invariants provides fairly detailed information about the world immediately, without much need for inference, because the relationships of the specific objects present provide unique information to our perceptual systems which cannot be duplicated except with identical or nearly identical objects. A little explanation on this will help explain why this relationship of variants and invariants—which relies on locomotion—provides direct information about the world.

If the stapler on my desk really were just a picture of a stapler, when I stand it would change very differently than if it were a real, three-dimensional stapler. When it is a two-dimensional fake, the relationship is dissimilar, and unique to the fake image and the desk. If I stand, the two-dimensional image does not change in the same way a three-dimensional stapler does. There will be no variants. All parts that were visible to me when sitting will still be visible when standing, with no new aspects or angles of the stapler revealed. Thus, the input is not underdetermined. In fact, it's very determined because the interaction between variants and invariants is exclusive. Though the stapler does not change, other three-dimensional objects will change as I stand, revealing both their three-dimensionality and the image of the stapler's two-dimensionality. Considering this scene as a whole, only a desk of a certain size, color, and texture could reflect light the way the desk I'm perceiving does, and only an outlet of a certain size, shape, and location could interact with the desk in the way it appears to be doing so. In other words, once I start moving around there is only one possible set of objects that could look the way these objects look to me. Any change whatsoever in their size, shape, color, or texture will present a different set of variants and invariants and thus a different world to perceive directly.

If this model is accurate, two aspects of it offer significant challenges to the traditional model: (1) a lack of inference and (2) a dependence on sensorimotor activity. As mentioned before, Gibson's suggestion is that the information received by the retina is not underdetermined and therefore does not require one to fill-in-the-blanks regarding the continuity of objects as well as their relative size and location. Instead, the light received by the retina basically contains all the physical information about the objects perceived (size, shape, texture, location, relationship to other objects). However, because the size, shape, and position of objects can only be revealed through changes in the ambient optic array, one's ability to perceive any given set of objects is now tied to one's ability to explore and traverse the environment.

Initially, this may sound odd as one can imagine perceiving something like the interior of a grocery store—filled with myriad object of different sizes, shapes, and textures—without having to move an inch. Gibson's point is that it would be impossible to perceive the different objects in this environment as separate until one begins to move about. The deli section in the back is only revealed to be in the back based on how it moves (or doesn't) in relation to the aisles closer to you; the bananas are recognized as being behind the apples only when they are perceived as being blocked by them at certain angles. Without movement, and the changes in reflected light that reveal variances in the environment, the entire scene could be a two-dimensional photo where only one actual object is present in front of the eyes. If Gibson is correct, movement combined with the ability to receive visual input provides direct information about the environment without the need for inference or encoding for it to be cognitively useful. But if this is so, what kind of information is being perceived? His answer: affordances.

Affordances are not necessarily unique to human beings, but are exclusive to animals. It refers to what a given environment *affords* or offers an animal. The desk, for example, offers a place for me to rest my hands or head—it's flat, extended, seemingly supported surface has enough space for my hands, doesn't appear harmful, and looks as if it won't collapse. A tray of broken glass, on the other hand, doesn't afford the same perception—it is not perceived as a good place to rest my hands due to its uneven, sharp surface(s). Thus, affordances can be perceived as advantageous or disadvantageous. The objects in the environment, and the environment as a whole, can have positive or negative affordances in them. One of the more important and intriguing aspects of this is that these affordances are not objective; they are subjectively perceived by the animal depending on its form and abilities. In this way, a human with opposable thumbs might perceive a pen as an object that affords the ability to write, twirl, throw, stab with, etc., whereas the pen does not have the same affordances for a snake.

> "The affordances of the environment are what it offers the animal, what it provides or furnishes…it implies the complementarity of the animal and the environment…they have to measure relative to the animal. They are unique for that animal." (Gibson: 127)

As described so far, affordances could be fully compatible with the traditional model of cognition and perception, with inference allowing one to determine what possibilities the world affords. That is, one could see a pen for the first time and, based on awareness of one's own dexterity, make certain inferences regarding the pen's potential use and usefulness as a handheld object. However, Gibson is clearly against this and is explicit in claiming they are perceived without inference. "This is a radical hypothesis, for it implies that the 'values' and 'meanings' of things in the environment can be directly perceived." (Gibson: 127) By 'value' and 'meaning' he is referring to the physical opportunities or obstacles the objects in the environment have in relation to the specific body one has. When hiking through the woods, a stump can be valued as a place of rest because its physical components are suited so that a human being could sit on it, giving it a meaning comparable to "seat." But the word "comparable" is very important here. Gibson is not suggesting that *classifications* of objects are perceived directly. That is, one does not immediately perceive "pens," "staplers," "watches," or any other arbitrarily named classifications that help us group and arrange concepts. As he says, "you do not have to classify and label things in order to perceive what they afford." (Gibson: 134) Instead, one immediately takes in certain features of objects based on one's potential interaction with it. It is possible to immediately perceive if an object is graspable because, according to Gibson, one can know the size of one's own hand without reflection or inference. Likewise, one can perceive if it is sharp, flat, relatively solid, and so on. These aspects are what provide affordances, and it is not necessary to know all the features of an object and have it properly categorized in order to perceive them. Taken alone, these are not very robust values or meanings, but they are the basis for much more complex concepts and categories communicated with terms like "banana," "desk," "stapler," and "penny." And whereas these last terms are based on language and context, affordances are much broader and are recognized due to evolutionary history.

As species evolve in certain environments, the perceptual systems of these species adapt to recognize different features of the environment as beneficial and detrimental. Also, since the affordances are species dependent, it helps explain why some objects may be perceived to offer a certain affordance to one species and not another, even when the species share the same environment. A human, therefore, can perceive an apple as graspable whereas a dog may not. This is not due to the human's ability to make an inference about the apple, but because the human's evolutionary history has attuned its perceptual systems such that it immediately recognizes the apple as complementary to its form and goals. This may sound similar to the idea of "coupling" but there is a subtle yet significant difference between Gibson and most contemporary supporters of embodiment. Though affordances are due to one's unique evolutionary history and embodied form, he doesn't say the meanings of objects are bound within coupled relationships between mind, body, and world. Instead, as he supports a notion of direct perception, he holds the value and meaning of affordances to be external to the perceiver. Thus, Gibson seems to be much more amenable to the idea of content externalism than coupling.

For this reason, several philosophers, such as VTR, make a point to separate themselves from this aspect of Gibson's theory while still embracing much of his explanation of perception. As VTR state, "Gibsonians…attempt to build up the theory of perception almost entirely from the environment," while VTR's approach "proceeds by specifying the sensorimotor patterns that enable action to be perceptually guided, and so we build up…from the structural coupling of the animal." (VTR: 204). This latter approach seems to be the more common (and more controversial) approach taken by many philosophers who, though inspired by Gibson, are more influenced by VTR's theory of *enaction*. The explanation of color experience in the section on coupling gave an idea how enactive perception works via the coupling of mind, body, and environment. Rather than go into more detail on VTR's theory specifically, we can look to the work of Kevin O'Regan, Alva Noë, and the late Susan Hurley for an understanding of an enactive theory of perception that builds upon the work of Gibson and yet denies the independence of the environment in line with VTR.

Together and separately, O'Regan, Noë, and Hurley have agreed with Gibson's claim that action is essential for perception. They deny the traditional model's emphasis on the brain in favor of an approach that includes the bodies of perceivers as constituents of perception. This means there is no one-to-one correspondence between neural activities alone and visual experiences, and any brain-in-a-vat scenario in which the brain is imagined to have a visual experience similar to an embodied brain is impossible. Adopting a form of disjunctivism, they say neural activity may be necessary for perception and visual experience, but neural activity alone is not enough to produce visual experience. There may be some sort of experience by stimulating a brain in a vat, but it will be of a different kind than visual perception. Neural stimulation is not sufficient because vision and other perceptual systems and experiences are inextricably linked with the body in that a given environment is perceived in a certain way due to one's potential embodied interaction with it. Therefore, rather than viewing the perceptual and motor systems as separate systems that provide the central cognitive unit of the brain with information input and motor output, respectively, "perception and action [are] mutually and symmetrically interdependent." (Hurley, 2001: 31) In this way, an experience of vision is not constituted solely by neural representations that are activated due to interaction with the environment. Instead, the visual experience is constituted by what an agent and his/her body are doing.

> "If the content of perceptual experience depends crucially on the environment, as well as on skillful motor capacities and capacities for directed attention on the part of the perceiver as a situated agent in the environment, then it cannot be assumed without argument that there is any such thing as a minimal neural substrate sufficient to produce conscious experience. Rather, the substrates of consciousness—in particular of visual perceptual consciousness—seem to cut across the brain-body-world divisions." (Noë & Thompson, 2004: 25-26)

Such a strong claim is going to have many parts to it, but at its core is a slightly amended idea from Merleau-Ponty. In Merleau-Ponty's *Phenomenology of Perception*, there is a chapter entitled "The theory of the body is already a theory of perception," if 'the body' and 'perception' are switched you have the central thought of the enactive account: the theory of perception is already a theory of the body. Perception is both the implicit and explicit understanding of the effects of movement on the environment around us.

This can be understood by considering the experience of looking at a cube. When looking at a cube, the entire six sides, twelve edges, and eight vertices are never visible at one time from one vantage point (assuming no mirrors are involved). At most, three sides will be in view at one time. And as one moves around the cube the sides, edges, and vertices change dramatically. Some aspects will disappear while others will appear. Noë refers to this as the “visual potential” of the cube. Whether due to one’s change in position or a change in position of the cube, the cube has the potential to reveal the entirety of its sides, edges, and vertices to a perceiver, but only based on movement. “Any movement determines a set of changes in perceived aspect; any set of changes in perceived aspects determines equivalence classes of possible movements.” (Noë, 2004: 77) If there is not a change in position then there is no way to perceive the cube as having more than three sides. Only through movement is one able to perceive every aspect of the cube. Otherwise, the object may be classified as no more than three-sided, and as having unequal sides. The movement that reveals the visual potential of the object allows one to classify it as a cube. Thus, when one encounters the visual potential of a given object, one can encounter its actual shape. So to experience a cube as a three-dimensional object is to understand how its appearance will change as one moves. To relate this back to Gibson, we could say that the visual potential of an object is the set of possible variances included within the invariant substance of a given object.

Visual potential, according to Noë, is vitally important and ubiquitous to perception. Because of our limited vantage points, every object in the world has some aspect hidden from a perceiver. Even if an object is flat (such as a square compared to a cube), the understanding that it is flat is based on its back side not extending far, if at all, from the front side—though it is hidden from view until the perceiver explores the visual potential of the object. However, this is still aligned with the Gibsonian claim that affordances are external to the perceiver. What the cube (or any other object) means to an agent could still be externally based and revealed through, or even synonymous with, its visual potential. In response to this, supporters of the enactive approach stress that affordances are only valuable insofar as they are defined by ways in which one could interact with objects in the world. This interaction is defined by the unique body a perceiver has and the ways in which s/he can maneuver her/his body; the myriad possible ways in which one can interact with a given object or environment can be referred to as sensorimotor contingencies. And the linguistic coupling of "sensory" and "motor" in this last term is indicative of the coupling between the body and perception.

When perceiving, one doesn't first sense the world and then determine what motor activities are appropriate or possible via a rational process. Instead, the things one senses are based on and defined by the body one has. Thus, an object can be perceived as graspable because one has hands with which to grasp; an object is perceived as edible (or *possibly* edible) because its shape and/or texture are amenable to one's alimentary canal; and so on. The difference here between Gibson and enaction is that enaction considers visual experience itself to be metaphysically constituted by sensorimotor contingency—it's not just that the world is perceived through doing, perception itself is based on what one can do in the world. "An analogy can be made with 'life': life is not something which is generated by some special organ in biological systems. Life is a capacity that living systems possess. An organism is alive when it has the potential to do certain things." (O'Regan, Myin, & Noë, 2004: 79) Likewise, perception is not generated by the perceptual systems or neuron activations alone; perception is something we do as embodied organisms.

The idea that visual experience is itself a sensorimotor contingency is centered on the claim that experience creates tacit expectations regarding interaction with and understanding of the external environment. Sensorimotor contingencies are based on the "structure of rules" governing each sensory modality. For example, a change in light is expected to change the visual perception of objects, but not the auditory perception. The rules governing visual perception will have to include retinal shifts and flow patterns, the effects of head movements and blinks, and much more regarding the visual system, whereas the rules governing auditory perception will not be affected by retinal movements or blinks. Additionally, these contingencies come in two varieties: those due to the physical capabilities of the sensory system and those due to the character of objects. In vision some are based on, say, turning one's head or squinting one's eyes in order to manipulate positioning and focus, whereas others might be based on placing an object in sunlight to reveal its color(s). These rules exist but are not innately known to a perceiver nor do they need to be expressed propositionally in order to be understood. They are revealed and grasped through sensorimotor probing of the world. By moving through an environment and experiencing changes in light, changes in position, changes in vantage points, etc., one begins to recognize sensorimotor contingencies (and the corresponding rules) implicitly, not propositionally.

This is why/how a toddler learns to play peek-a-boo. The child will begin to understand that putting her hands over her eyes blocks another person from view and only when her hands are moved away is the other person visible again. Once the child is able to play the game correctly, she is revealing awareness and skill of certain sensorimotor contingencies. And, through continued exercise of these abilities, the child (and/or the child's brain) will become attuned to the laws of sensorimotor contingencies. This might be demonstrated at a later time when she wants to block something from view, such as a bright light, and (seemingly) instinctively covers her eyes with her hands. There is an understanding evident in her actions as to how one can block something from view; however it doesn't need to be consciously explicit to her nor based on a rational, propositional process. Instead, most of it is implicit and practical.

The knowledge of sensorimotor contingencies—even though the brain likely plays an essential role, because it may be necessary for "tuning"—is not of the traditional computational sort, nor is it exactly like Gibson's ecological perception. O'Regan and Noë say visual experience does not occur unless the perceiver is exploring the environment in a manner that is governed by the sensorimotor contingencies of the visual system and those fixed by the character of the object(s), and the perceiver must be "tuned to" these sensorimotor contingencies. (O'Regan & Noë, 2001: 943). By "tuning" they are referring to the manner in which collections of nerve cells become wired, so to speak, to perceive certain frequencies. Over time, with continued exposure to similar input from the various sensory systems, receptors tune in to specific patterns of frequency waves and exhibit a pattern recognition response wherein they respond to familiar frequency patterns among the myriad input frequencies with which one is bombarded each day. And, by being attuned to the same patterns over and over again, a particular pattern or picture of reality is formed. Patterns that are unfamiliar will then often be ignored because they don't fall within one's usual receptor input. The result is that one's brain ends up filtering input from a limited set of perceptual patterns. Only if both of these conditions—sensorimotor contingency-governed exploration and tuning—are met does visual experience occur. In this way, the meaning perceived from objects cannot exist separate from the perceiver, as Gibson suggested, because the experience of visual perception is a coupled experience between one's body, the world, and the implicit understanding of potential interaction between the two. Meaning is not "out there" waiting to be perceived or pregiven in any sense; meaning is enacted via the skillful interplay between perceiver and environment.[7]

Another way of presenting the enactive approach has to do with modularity. The traditional model presents the processes of cognition as being *vertically* modular. (Fodor, 1983) That is, each module performs a function, generating or affecting a mental representation, and then moves the representation on to the next module. The visual module, for example, retrieves information about color, location, motion, etc., from the various parts of the physiological visual system that serve domain-specific purposes. This input is then combined to create a unified representation of a given object or scene that is then moved to the central module where perception and action interact. Since this is believed to be computational it is similar to parts of an equation wherein expressions within parenthesis are to be solved before they become factors in the larger equation. This equation, as a whole, is rational thought—cognitive processing is a linear path from perception to cognition to action, with the various perceptual modules providing specific types of input to be factored into the cognitive "equation." Many equations can be occurring at once, but each is believed to operate in this vertically modular way.

Hurley, on the other hand supported a *horizontal* modularity.

> "Rationality reconceived in horizontally modular terms is substantively related to the environment. It does not depend only on internal procedures that mediate between the input and output, either for the organism as a whole or for a vertically

[7] Some, including Noë, Hurley, O'Regan, Gallagher and others who have incorporated the phenomenological character of experience into their explanation of cognition, will say meaning relies on a phenomenal aspect of intentionality that partly constitutes it. Since the phenomenal character of a given experience (assuming such character exists) is unique to the individual, meanings cannot be pregiven or exist in any purely objective manner. They are necessarily subjective and intertwined with the embodied individual.

> bounded central cognitive module. Rather, it depends on complex relationships between dedicated, world-involving layers that monitor and respond to specific aspects of the natural and social environment and of the neural network, and register feedback from responses…Very crudely, some layers get turned on and others turned off, in a totality of ways that count as rational overall in the circumstances. On this view, rationality is a higher-order property of complex patterns of response, which emerges from the layers of direct dynamic couplings between organisms and their structured environments." (Hurley, 2001: 10)

Accordingly, cognition does not operate by different modules providing specified domains of information that the central module then processes before determining a motor output. Instead, input from perceptual modules overlap, and any resulting mental representations are processed with motor action already built into them (or encoded, to stay with computational terms). This is why, when an object is perceived by the visual module to be moving rapidly in the direction of one's face—say a stray soccer ball—, a person likely shuts his eyes and/or moves his head without rational deliberation. Due to the skillful probing described earlier, an agent who has met the criteria for visual experience will have certain reactions to the environment that reveal coupling of the modules.[8] A man who has a soccer ball flying at him does not receive the visual input that an object is rapidly approaching, which is then processed with information from other modules—such as proprioceptors, the peripheral auditory system, etc.—which provide a total amount of data, which is then mediated via internal procedures that allow him to cognitively form conclusions about the speed at which the ball might be traveling, where it might hit him, how soon the impact will occur, and how he ought to move so as to either avoid it or protect himself. If this were the case, perception could be thought of as a means to action (and action as a means to perception), wherein each is instrumentally valuable to the other. Hurley accepts that

[8] This could be countered by the claim that experience has allowed the cognitive "equation" to be processed so rapidly that it only appears to be coupled, or occurring without rationality. An idea similar to this was discussed in Raftapoulos (2009) regarding the mind's ability to individuate objects at a non-conceptual, non-conscious level as quickly as 50ms after stimulus (p. 87).

perception and action are instrumentally interdependent, but she also claims they are constitutively interdependent.

Cognition, then, is comprised of a network of interrelated subpersonal layers, each of which is a "complete input-output-input loop, essentially continuous and dynamic, involving external as well as internal feedback." (Hurley, 2001: 8) Sensory and motor processes are coupled, as well as the neural network of an agent with his/her environment (because the potential and past action of the agent is critical to the feedback loop). If this is accurate, there is perception/action content rather than separate perception and separate action content. The content is generated by "leaky," intertwined modules which prevent the separation of action from input and allow the man to duck or cover his face when a ball is flying at him via a process that is neither linear nor merely instrumental; they are coupled to the point of being inseparable content that is both determined by and determines how he will move and what he will perceive visually.

What should be taken from all of this in regards to the traditional model of cognition? Does this mean the traditional model is in need of replacement, or merely in need of amendment? The enactive approach is clearly aimed at distinguishing itself from the traditional model and, at least according to Noë, in a manner that demands more than just amendment. He claims there are two main implications of the enactive approach (Noë, 2004):

1. Only a creature with certain kinds of bodily skills—for example, a basic familiarity with the sensory effects of eye or hand movements, and so forth—could be a perceiver (because perceiving is a kind of skillful activity that requires perceptual modes of self-awareness).

2. We ought to reject the idea that perception is a process in the brain whereby the perceptual system constructs an internal representation of the world. Perception is a kind of skillful activity on the part of the animal as a whole.

(1) Entails the impossibility of brain-in-a-vat scenarios or any hope of cognition based on symbolic manipulation alone. (2) Denies that there is any sort of neural correlate to consciousness that accounts for perceptual experience. If tied in with Hurley's description of horizontal modularity[9], this represents an upheaval of several tenets of the traditional model to the point that it offers if not a wholesale replacement, at least a new version of cognition that alters many of the core ideas supporting the traditional account.

However, before adopting such a stance, much more needs to be discussed. As will be seen in chapter 4, significant challenges arise regarding whether the role between action and perception, or mind, body, and world are actually metaphysically linked in such a way that cognition cannot, in any possible world, function apart from mind-body-world interaction. If they are metaphysically dependent in that way, Hurley, Noë, and O'Regan might be offering an account that should displace traditional cognitive science. But if that strong metaphysical standard is not necessary for cognition, these theories of perception may only offer new explanations for how sensorimotor information is encoded in perception, but do little to reorganize the mental sandwich.

[9] Hurley is not alone in this. It has been suggested and supported by many, including Clark (1997), Thelen and Smith (1994), and Brooks (1991).

3.3 Representations are not amodal

According to the traditional model, mental representations are physical yet purely symbolic. As such, even though I may think about watching someone score a bicycle kick in the finals of the World Cup, the ball, the goal, the field, the sounds, etc., are not really there in some miniature version somewhere in the brain. Nor is there a little version of someone, complete with head, arms, torso, legs, and feet, in my head completing the bicycle kick as I'm thinking about it. All that is actually present in the mind are combinations of mental symbols which, when put together in a certain way, result in my ability to imagine the entire scenario of a player scoring the goal. Analogously, just as the 1s and 0s of binary code—and the various ways in which they can be combined—do not depend on the particular screen, ports, or keyboard a computer has, the mental representations which human beings possess—and, more importantly, the symbols which constitute them—do not necessarily depend on specific appendages and/or sensory systems a human being possesses. All that truly matters is the central processing unit's ability to compute and the symbolic data that makes up the computations.

If this is accurate, I should still be able to think about someone scoring a bicycle kick goal even if I am reduced to nothing more than a brain in a vat. So long as certain neurons are receiving electrical stimulations that allow me conceptualize the stadium, the player, the sounds, the grass, etc., it doesn't matter if those mental representations are perceived from first-personal, embodied experiences in the world or from very specific stimulations to the brain. My thoughts, even though they may be about my embodied interaction with the world (being in the stadium and watching the kick), are nothing more than a combination of formal mental symbols of encoded semantic content. This combination of symbols and representations constitutes the mental state and, assuming they can function independent of a sensory system's interaction with the world, it does not matter whether the neural stimulation is coming via visual, olfactory, or aural contact with the world in the ways human eyes, noses, and ears general work. The embodiment doesn't matter for the cognitive capability, only the neural stimulations necessary to cause the activation and use of mental representations matter.

When considering cognition as a whole, all the usual modes of input—each of which relies upon different sensory cortices to represent different features of the world—only matter in terms of how they are symbolically encoded and translated. Even though a smell will be encoded differently and via different processes than a sight or a sound, that encoding could be replicated (hypothetically) by anything that can manage to manipulate neurons in the same way in order to cause mental symbols to mean that smell. Thus, the semantic content of representations in the brain, even if that content is modal sensory information, is not continually dependent upon the sense modalities that provide(d) the initial input. It is metaphysically necessary that something provide the semantic content, but once provided, the object and sensory system through which it was perceived are not necessary for each mental token with that semantic content to continue to exist. Though they are causally linked, mental representations are distinct from and operate independently of sensorimotor input and output. In this way, thoughts "must be couched in an amodal medium, carried out in a central processing system that functions independently of input systems." (Prinz, 2002: 151)

This reference to a medium is significant. When discussing amodal symbols from the perspective or traditional cognitive science, the amodality refers to the structure and syntax of the symbol within cognition, not its semantic content. Using binary code as an example again, the 1s and 0s have a structure and syntax that function independent of the ports, screen, or keyboard of a computer. Certain combinations of 1s and 0s will have specified meanings that will affect what is shown on the screen or how keystrokes are interpreted, but the code itself is carried out in a medium separate from the input/output systems. The code is amodal in composition and function even if the semantic content it carries is modal.

Embodiment theories generally deny this sort of amodality. Instead, many theories claim that all mental representations are inherently modal in that the input systems are continuously coupled with (at least) every sensual representation so that any mental experiences of, say, a sound is possible only when parts of the auditory system are activated. However, this is a strong claim with a great many variations and a good deal of explanation needed.

Does it mean that the parts of the sensory systems that *can* interact with the external world *must* interact with the external world in order for mental representations of tastes, sights, sounds, and the like to be generated? Yes, but this is only a claim of empiricism regarding sensual content of concepts. To say one cannot have a concept of the taste of honey without having tasted honey at some point is not exclusive to embodiment. Traditional cognitive science endorses this claim as well by saying sensory stimulations play a necessary causal role in the formation of the symbolic mental representation of the taste of honey. That is, one must taste honey to have any mental content of honey's taste, yet interactions like this with honey only generate modal information that will become part of the HONEY concept *type*. Once the information is processed and translated, the concept is believed to no longer rely on the tongue or other sensory systems that provided the sensual information for the concept for it to be used cognitively. The translated concept is now amodal, even though it includes symbols with modal semantic content, and each subsequent token of HONEY can exist and function amodally as well.

Since embodiment theorists are at odds with traditional cognitive science, do they want to say sensory systems must continue to interact with the external world in order for mental representations of tastes, sights, sounds, and the like to be used cognitively? In other words, are mental representations so tightly coupled to sensory systems and bodily interaction with the world that one must not only have tasted honey at some point but must also continue to have the ability to potentially taste (or perceive in some way) more honey in order to continue to utilize concepts of honey-taste in cognition? The answer is vague: somewhat yes and somewhat no, depending on the theorist; but mostly no. If the above position is endorsed, it could entail that if one loses the ability to taste then taste concepts will be lost as well. This is an extremely strong claim that is not widely endorsed due to plenty of evidence of individuals losing certain senses due to bodily damage, such as blindness, yet still retaining memories with information pertaining to the lost senses. However, this claim also depends on how a theory chooses to delineate the bounds of a given sensory system.

If a sensory system such as vision is defined as the non-neural systems and processes—basically from the optic nerve outward—visual mental representations do not depend on the sensory system to continue to exist or function cognitively, because blindness does not entail a loss of visual thoughts. But if a sensory system is defined as the neural *and* non-neural—bodily subsystems and processes—this may not be the case because damage to the neural pathways and processes responsible for certain perceptual experiences can result in an individual losing the ability to apply modal concepts to non-modal concepts.[10] Later in this section a case will be discussed in which subjects with damage to the mesial temporal region of the left hemisphere

---

[10] This assumes a modularity of the mind wherein certain parts of the brain are independently responsible for the generation, experience, and storage of different types of mental representations. This is an acceptable premise for traditional cognitive science as well as embodiment theories, so I have chosen not to go into detail about any debates surrounding it.

lost the ability to apply specific colors to object concepts. But even this doesn't definitively prove that one loses modal mental content without access to active sensory systems. At most it seems to indicate a problem in knowledge rather than cognition, which will be discussed near the end of this section.

With this in mind, the next question should be if sensory system modality includes neural and non-neural systems and processes. If not, and sensory systems begin or end just outside the brain, aside from the implications on retaining mental representations with modal content, nearly all sensorimotor mental representations would exist amodally and likely continue to function without actual or potential contact with the external world. They are amodal in the sense that, once formed, they retain their semantic properties regardless of the state of the sensory system responsible for the initial input. Though the semantic properties of the representations may be attainable only from information provided from a certain sensory system and from neural processes in a specific area of the brain, the sensorimotor systems only play a causal role and are not necessary to sustain the existence and function of the representation. However, saying sensory systems begin and end outside the brain is not a widely endorsed position by traditional cognitive scientists or embodiment theorists—the majority in both camps claim that some neural processes must be considered part of the sensory system along with the other nerves, tissues, and processes that connect the brain to the world. For example, the visual perceptual system is generally divided into five layers of organization (reception, transduction, transmission, organization, and interpretation) that reaches from the retina to the primary visual cortex and accounts for receiving, translating, and interpreting all visual information. On the traditional model, the output of the primary visual cortex, after receiving input from the external world, is modal semantic content in an amodal language of thought that can be used independent of the sensory system—which is why blindness doesn't result in the loss of visual memories and one can still have (seemingly) visual experiences when dreaming.

This is the crux of the dispute between traditional and embodied accounts in regards to mental representations: embodiment theories argue that the successful cognitive function of modally-specific semantic content depends on that representation's continual access to and with the neural (and sometimes bodily, non-neural) processes necessary for the initial organization and interpretation of the external input. According to embodiment, at no point does a representation with modal semantic content exist independent of the sensory system that generated it—the representation itself always remains modal. Any time there is a perceptual experience based on external stimuli *and* when one is imagining, remembering, or thinking about sensorimotor-based concepts, one is only able to achieve those feats of cognition and utilize such mental representations due to distinctly modal activity in the brain, in terms of both function and content. The modal semantic content is constituted by modal neural processes and is thus continually coupled to them in order to function cognitively, rather than merely being caused by them. If this is accurate, then modal neural processes constitute representations (which are invariably connected to the variegated, specialized forms of sensory input provided by the different sensory systems). And any mental representation with sensorimotor content will never achieve the amodal operation called for in the traditional model.[11]

One way to consider a "modal" symbolic construct, as opposed to an amodal one, is to think of symbols as analog rather than formal. Lawrence Barsalou described symbols derived

---

[11] For now, I'm discussing sensorimotor modal content involving taste, touch, smell, sight, or sound because this is the content most amenable to embodiment theories as it relies most clearly on embodied experience. One might wonder how more abstract amodal content might fit in here, such as mental representations involving CALCULUS or GENUS. Some of this will be covered later in this section and even more will be discussed in the next section on language.

from embodied experience and retaining information about sensorimotor processes as "perceptual symbols." At the core, perceptual symbols are modal and analogical. The symbols are modal because they are not represented in a system or area of the brain independent of the perceptual system(s) responsible for their creation. Instead, the symbols seem to be represented in the very same perceptual systems that produced them. (Empirical support for this will be discussed later in this and the next section.) "On this view, a common representational system underlies perception and cognition, not independent systems." (Barsalou, 1999: 578) And, due to this, the symbols are also analogical. The activation in perceptual systems when accessing or utilizing representations of concrete objects and other sensory concepts can be interpreted as a sort of simulation of the original perception, thus making the semantic content of the representation analogous to act of perception. Whether communicating to another person or attempting to remember a given concept, "the structure of a perceptual symbol corresponds, at least somewhat, to the perceptual state that produced it." (Barsalou, 1999: 578)

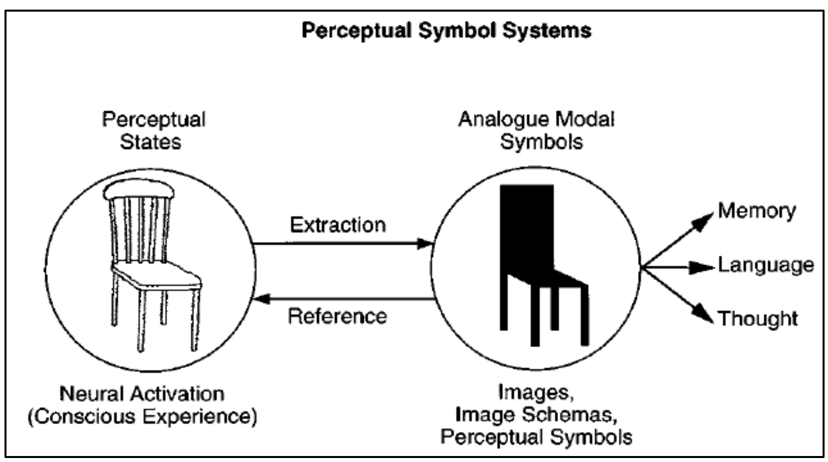

Figure 3.1 Barsalou's Perceptual Symbol System

The figure to the right (from Barsalou, 1999) helps explain the operation of perceptual symbols. When an object, such as a chair, is perceived, modal properties are extracted from the initial perceptual state in order to construct the images, image schemas, and perceptual symbols that will provide the semantic content of the representation of the chair in memory, language, and thought. The properties remain modal (and analogue) because they rely upon the original neural activations in order to provide and maintain their semantic information. By analog, Barsalou is not endorsing a form of imagism wherein one experiences picture-like representations in the mind, as supposed by the early empiricists. As he explicitly states, "perceptual symbols are not like physical pictures; nor are they mental images or any other form of conscious subjective experience…instead, they are records of the neural states that underlie perception." (Barsalou, 1999: 582) Thus, when one thinks of the concept CHAIR, it is not as if one is must have a mental image of the chair sitting on the stage that is the mind's eye, to use an illustration from Hume. When one thinks of CHAIR, some of the neural processes involved in perceiving a chair reactivate. Though this could create a sort of conscious experience of imagining the visual image of a chair—which would seem to indicate imagism—Barsalou believes perceptual systems also function unconsciously. And these unconscious neural representations, as opposed to conscious mental images, "constitute the core content of perceptual symbols." (Barsalou, 1999: 583)

If Barsalou is correct then, contrary to traditional brain-in-a-vat scenarios, the belief is that only physical sensory systems like the ones of the human body can provide the sort of input necessary to allow our neural subsystems to generate mental representations with the modal content they have. Though the brain may be a bunch of electrically stimulated neurons, only stimulations tied to the elaborate, specific sensory systems of the human body can generate the types of thoughts humans have. Thus, in spite of the fact that a person can lose the ability to see, yet retain visual memories, an embodiment theorist could claim that such memories rely on the parts of brain previously involved in processing visual perceptions being stimulated in order to simulate visual perception. If a person goes blind no new input will be available to the neural areas that previously processed visual input, but those portions of the brain could still be active and allow for a blind person to continue having visual thoughts—the portion of the sensory system that still works may constitute visual memories and thoughts. Should that be damaged, perhaps the subject would never have visual thoughts again. If so, then concepts and mental representations with modal semantic content never achieve the amodality supposed by the traditional model because their function is reliant upon sensorimotor processes.

The question now ought to be "Why should anyone believe this account more than traditional cognitive science?" And this is where embodiment theorists are combining neuroscientific findings with their philosophical argumentation. Below are the three main reasons, philosophical and scientific, for concern over amodal representations:

1. *There is little empirical evidence for amodal symbols.*
   Though they offer a wonderful account for formalizing cognition and (at least initially) hope for artificial intelligence, there is a lack of clear evidence in neuroscience to support their existence.

2. *The grounding problem.*
   If one is to accept a thoroughly amodal account of cognition, it is difficult to explain how cognition interacts with perception and action because the central processing unit of traditional cognitive science lacks necessary ties to either.

3. *"Traditional theories increasingly face a lack of understanding about where the brain stores amodal symbols and about how amodal symbols could be consistent with neural principles of computation." (Barsalou, 2008: 620).*
   A great deal of research has attempted to map the activity of the brain while a subject performs various cognitive activities and has shown that most mental activity—even when a subject is not performing a sensory- or motor-dependent task—correlates with parts of the brain tied directly to sensorimotor processing. If cognition is comprehensively amodal, it's possible the amodal symbols and representations that constitute one's thoughts are stored in parts of the brain not related to sensory or motor processing. Thus, if traditional cognitive science is correct, when one imagines scoring a goal from a bicycle kick it should be possible that the cognitive activity not require the activation of neurons generally tied to the motor abilities required to perform an overhead kick. However, this doesn't seem to be the case. Very often, even if one is only imagining a physical action, specific areas of the brain required to complete that activity in the physical world become activated. This leaves the door open for challengers of the traditional model to ask where the amodal symbols reside since much of the brain activity appears to be fundamentally modal.

Though these three claims can be considered separately, the evidence supporting them cannot. Because of this, the rest of this section will look at separate studies and findings and then explain how they support the claims above. The discussion will begin with a look at studies in perception-action coordination, followed by studies of memory and conceptual processing. The goal is to gain a clearer understanding of how cognition actually appears to operate in terms of modality, rather than just how amodal and embodiment theorists might like it to be. It is important to note the core of this section is not about denying amodality entirely. Because traditional models like Fodor's generally take the hardline stance that *all* mental representations are amodal, one does not need to make the case that *all* mental representations are modal in order to falsify it. All that is needed is proof or valid argumentation that at least *some* representations are modal in the sense that either the semantic content they possess is continually dependent upon the activation of distinctly modal neural processes and/or the representations themselves can never be considered separate from modal neural processes.

"Some" modality in mental representations may sound like a claim far too weak for full-fledged embodiment. It is. Much of what will be discussed in this section falls under that banner of what is generally called "grounded cognition." Grounded cognition theorists basically claim that cognition is grounded in sensory and motor processes (thus answering #2 from above). This does not entail all semantic content is modal, but it does entail all mental representations—some of which may include amodal concepts—must ultimately be grounded in sensory and motor modalities. This also delineates some grounded cognitive theorists, such as Barsalou, from theorists such as O'Regan, Noë, and Hurley. Where the latter claim the parts of a perceiver's body that come into direct contact with the external world are constituents of perception, Barsalou and others claim the non-neural sensorimotor parts of the body play an essential causal role in the formation of mental representations, though not necessarily a metaphysically constitutive one. But before the metaphysical issue of constitutivity is discussed in chapter 4, it can at least be stated that they all agree with the claim that all semantic content may be derived from embodied activity. The primary reason for this claim is evidence that nearly any thought seems to rely on reactivating sensorimotor neural processes. The basic idea is that a thought of something like kicking a ball only occurs because one's mind is cognitively simulating the action by activating the neural processes necessary to actually kick a ball. These finding will be covered below, and though they are aligned with O'Regan, Noë, and Hurley, Barsalou allows for a very significant loophole that mental content qua content may eventually be able to exist in an amodal medium. In chapter 4 I'll discuss how this may allow for grounded cognition and traditional approaches to be largely compatible. Yet, for now, the key feature of grounded cognition theories is the claim that all mental representations rely on embodied activity and sensorimotor processes in order to have any meaning or use. And proponents of grounded cognition believe this should be taken as strong evidence against the traditional model.

The effect of visual information on motor responses

This particular area of study is very close to what has already been discussed in the section on perception as perceptually-guided action. But we will now look at how this relates to mental representations by focusing more on the psychological and neurological studies of perception rather than the phenomenological and/or purely philosophical. In regards to perception-guided coordination this means looking at a subject's response to certain sensory stimuli (usually visual). The common finding has been activity that seems to reveal very specific modal responses. For example, when a subject observes an object, such as a cup, the neural activity is similar to that required for certain motor responses, like grasping the cup with one's hand. As Barsalou describes it, "as people perceive visual objects, simulations of potential actions become active in preparation for situated action." (Barsalou, 2008: 624) The simulations he is describing are reproductions of embodied activity that occur at a non-reflective level. That is, one doesn't have to imagine oneself grasping the cup as a conscious thought. Barsalou and others are claiming the perception of the cup, from the initial perception all the way through the stored properties of the mental representation CUP, is tied to potential embodied interaction along the lines of Gibsonian affordances. But, rather than simply positing this as a way of interpreting the activity of perception from the philosophical armchair, so to speak, advocates of grounded and embodied cognition are finding more and more empirical evidence that suggests sensory and motor modalities are inseparably fundamental to the *neural* process of perception as well.

To illustrate this point, consider a series of experiments conducted by Mike Tucker and Rob Ellis (1998). Initially, these should sound compatible with and supportive of the notion of affordances discussed in the last section. Yet they also provide a nice bridge into the more robust aspects of a modal theory of mental representations. Each of the experiments is designed to test if and how an object's relation to one's body affects how one represents it.

In the first of three experiments, Tucker and Ellis wanted to see how perceptions may be connected with motor responses. They weren't interested in more obvious things like if seeing a graspable object would trigger activation in the parts of the brain responsible for hand movements. They went a trickier route trying to see if angling a graspable object toward one of a subject's hands would prime the subject for actions with that hand. To this end, they showed subjects a series of objects and asked them to use a keypad—one on their left or right hand—to indicate whether each object was upright or inverted. The images were of commonly graspable objects (such as a saucepan, knife, and teakettle) some upright and some inverted, with the handles of the objects angled toward either the right or left hand of the subjects. The goal was to determine if the angle of the handle would affect the time it took subjects to establish whether the object was upright or inverted. Their hypothesis was that objects would be categorized quicker if the angle of the object were directed at the same hand needed to give the correct answer. For example, upright objects angled to the right would be categorized quicker if the "upright" keypad were on the subject's right hand as opposed to his/her left hand. The rationale behind this being,

> If the representation of a visual object includes action components, such as the preferential activation of the hand most suited to perform a reach-and-grasp movement, then one might expect this activation to facilitate simple keypress responses carried out by the congruent hand and, conversely, to interfere with those same responses carried out by the incongruent hand. (Tucker & Ellis, 1998: 833)

During the experiment both of the subject's hands were placed on separate response indicators, with the index finger on each hand resting on response buttons. One response button indicated "upright" and the button on the opposite hand indicated "inverted." The button placement would change from subject to subject, so sometimes the left index finger might be responsible for pushing the "upright" button and the right index finger responsible for pushing the "inverted" button, and sometimes vice versa. Subjects were then asked to respond as quickly and accurately as possible, pushing only the "upright" button if the object appeared to be upright and only the "inverted" button if the object appeared to be inverted. Tucker and Ellis found that response times were consistently faster and more accurate when the button indicating the correct vertical orientation aligned with the horizontal orientation of the handle. Thus, when subjects were shown, say, an image of an inverted saucepan with the handle angled to the left, their response times were faster by an average of 17 milliseconds, and there were fewer errors, when the "inverted" button was also on the left. Though Tucker and Ellis admit there could be myriad explanations for this occurrence, they believe the evidence "supports the proposal that certain action-related information—in this case the hand most suited to grasp the object—is represented automatically when the object is viewed in the peripersonal space." (Tucker & Ellis, 1998: 836)

The second experiment followed a similar method as the first, except that the "inverted" and "upright" response buttons from the first experiment were now both on the same keypad. The keypad was placed only on subjects' right hands and they were allowed to use only their index finger to answer. The goal of this setup was to provide a means to compare the supposed automatic action-related encoding effect observed in the first experiment. Oddly, the results of this setup showed subjects as having faster response times (whether upright or inverted) to objects oriented to the left. Though this might seem to go against the conclusions from the first experiment, Tucker and Ellis believe this to be compatible (though not corroborating) with the previous results. They believe the reason for the consistent left-right biases in the first were due to interactions between the objects' affordances and the left-right motor capabilities of the subjects. When a subject is restricted to using only his/her right hand, such affordance relationships cannot reveal themselves as clearly since the subject is engaging only one side of his/her motor potentialities in providing the response. As they say, "rather than object orientation automatically generating a left-right code by virtue of the visual properties of the object…it is the affordance for grasping by a particular hand that gives rise to the binary left-right distinction." (Tucker & Ellis, 1998: 838) Motor activations of the left or right hand, triggered by the object's orientation, will not be observable when only the right hand is involved in the response.

Finally, in the third experiment, they introduced a wrist-rotation component into the setup. Whereas before participants depressed a button to indicate an answer, this final setup required a clockwise or counterclockwise rotation of the right wrist to indicate whether an object was upright or inverted. Additionally, several new objects were introduced—such as a wine bottle, candlestick holder, and glue bottle—which lacked the graspable handle of objects in the previous two experiments but introduced a new element in that they all require a clockwise rotation of the wrist in order to be grasped (or counterclockwise if the objects were inverted). This experiment yielded results similar to the first experiment. The subjects' responses were slightly faster when the correct response rotation of the wrist matched the rotation of the wrist necessary for grasping the object[12]. Though many questions remain unexplored in these studies, Tucker and Ellis concluded these findings provide strong evidence that mental representations are intrinsically endowed with action affordances to the perceiver.

> According to this position, representing visual information involves representing information about possible actions and thereby potentiating them. One consequence of this is that intended actions are formed from, and informed by, already existing visuomotor representations. Actual actions are produced by the selection and elaboration of such representations. (Tucker & Ellis, 1998: 844)

Otherwise, the response times would not be affected by irrelevant visual information such as orientation of the object in relation to the perceiver's ability to grasp it. If valid, this conclusion supports the claims that there is little evidence for amodal symbols because the representation(s)

[12] The overall response times in this experiment were much longer than the first experiment, and the gap was less between responses where the correct answer required rotation required for grasping versus those where the answer required the opposite rotation.

utilized were imbued with modal information, and that very information allows certain aspects of cognition to be grounded in sensorimotor experience.[13]

To explain how such possible inextricable modal encoding could occur (and why it is at odds with traditional accounts of cognition), a closer look at brain physiology is required. Along the lines of the modularity of mind proposed by Fodor, a good deal of research has supported the notion that specific parts of the brain have specific functions and, in turn, certain functions of the brain have specific roles in cognition. One such example is the two-stream hypothesis of vision, proposed by Melvyn Goodale and David Milner (1992). According to Goodale and Milner, in addition to the visual input system which transfers information about the external world to the optic nerve, visual processing within the brain consists of two distinct pathways: ventral and dorsal. The ventral pathway travels from the primary visual cortex in the occipital lobe to the temporal lobe and accounts for object recognition and pattern discrimination; the dorsal pathway travels from the primary visual cortex to the parietal lobe and accounts for spatial processing. Basically, the dorsal provides the “where” and “how” in visual perception and the ventral provides the “what.” The ventral allows one to generate a representation of the world and the dorsal allows one to effectively determine how to interact with it. Together, these two streams account for everything a person visually perceives and how one is able to utilize that information to determine action.

The problem, when considering findings like those of Tucker and Ellis, is in the supposed independence of the streams. If the two streams are truly independent then irrelevant information—such as the direction of the handle on the saucepan—should not affect one’s ability

[13] This is not without controversy. It could be that visuomotor representations are involved in orientation representation in that amodal orientation representations cause visuomotor representations to be activated, and such a “grounding” relationship between the modal and amodal representations may be nomologically causal rather than metaphysically constitutive.

to determine the orientation of the object. This is because the orientation is determined via the ventral stream, whereas the direction of the handle and its relation to one's grasping ability is determined via the dorsal stream. Though the response variances in the Tucker and Ellis experiments were mere milliseconds when the handle direction and correct response side aligned, the consistency with which it occurred led Tucker and Ellis, as well as many others (Grezes, et al, 2003; Vainio, et al, 2008; Symes, et al, 2008; Tipper, Paul, and Hayes, 2006), to conclude the streams as unable (at least in some cases) to provide separable paths of processing. Just as Hurley claimed, there may be parts of the brain that have unique functions, but the modularity is not as clear cut as once supposed; how one perceives, judges, categorizes, and represents the world is affected by how one can potentially interact with it.

Reactivation of sensory-specific information during memory retrieval

Another psychological and neurological study backing the idea that representations are composed of modal processes comes from Mark Wheeler, Steven Petersen, and Randy Buckner (2000). Rather than providing data on the effects of modal perceptual information on outward action, this experiment focused on the response of the brain when subjects were asked to remember specific actions. In the experiment, subjects studied a set of 20 pictures and 20 sounds. They were exposed to each separate image and sound for a few seconds each. A descriptive label accompanied every image and sound, and subjects were told to memorize the visual and auditory stimuli in order to complete a test that would be given afterwards (though they were not informed what exactly the test would be).

After the initial rounds of stimuli exposure, subjects underwent an fMRI scan and were given a perception test and recall test. For the perception test, subjects were shown the same stimuli/label pairings as in the earlier part of the experiment and were asked to indicate whether the stimulus was an image or a sound. For the recall test, subjects were shown only the labels and were asked to indicate whether the stimuli that went with the respective labels were images or sounds. The first "test" was just a priming tool to allow the researchers to define sensory-specific areas of the brain activated during perception. The second test allowed the researchers to see if similar areas of the brain would become activated when the subjects were retrieving memories of those perceptions.

The final results of the study showed a very strong correlation between the activation of sensory-specific areas of the brain for perceptual processing *and* memory retrieval. For example, the perception of images during the perception test triggered activity in the middle frontal gyrus on both the left and right sides. When asked to recall labels matched with images during the recall test, distinct areas within the middle frontal gyrus (near left fusiform gyrus and bilateral dorsal extrastriate visual regions) became activated as well. Similarly, the perception of sounds during the perception test showed greater activation of the posterior inferior frontal gyrus on the left, and anterior inferior frontal gyrus on the right. When asked to recall labels matched with sounds during the recall test, activity was monitored in the bilateral auditory cortex. With such striking overlap of neural activity, the researchers concluded, "these data demonstrate clearly that vivid retrieval of sensory-specific information can involve the reactivation of sensory processing regions, supporting a reactivation hypothesis." (Wheeler, et al, 2000: 11129) In other words, when one attempts to access a mental representation of a specific sight or sound, the brain reacts in a way similar (but not identical) as to when the initial sights and sounds were processed.

Proponents of embodied cognition and other anti-amodal variations of cognition have viewed these findings as evidence that the mind does not translate external sensory stimuli into a fully formal, amodal language of thought when constructing mental representations. Instead, they believe the brain is forming representations that are decidedly modal in content and function; representations that, when applied, reactivate the input process just enough to simulate a perceptual experience somewhat like that of the original perceptual experience.[14] Thus, when one remembers the visual image of an apple, the brain—rather than utilizing neural processes completely removed from sensorimotor activity—may go through a process similar to when one perceives a real, three-dimensional apple in order to generate the internal remembrance of what an apple looks like.

The Wheeler, Petersen, and Buckner experiment can be used to support all three of the claims mentioned earlier. In regards to the first, the memory retrieval activity mapping aligned closely with perceptual activity mapping, indicating a stronger correlation with modal neural activity rather than amodal. Granted, this is not evidence that amodal symbols do not also exist, but it does provide reason to suppose the representations utilized by participants in the experiment had modal, rather than fully amodal, properties. This relates closely to the third claim as well because the neural activations appeared in areas of the brain associated with perceptual processing. If the representations were amodal, then it would be reasonable (though not necessary) to see areas of the brain dissociated with such processing activated. For example, an amodal mental representation of the sound of my son's name should not require neural activity in Broca's area for me to think of the sound of "Colin." If the traditional model is correct, that thought could rely on content stored in areas of the brain wholly disassociated from the neural processes necessary for speech production and language perception. However, there is a high

[14] I am stressing "somewhat" because many embodiment theorists will take a disjunctivist approach, stressing that cognitive processes that occur separate from embodied engagement with the world (such as contemplation, dreams, etc.), though derived from embodied activity, are qualitatively different experiences than perception.

likelihood that if I were wired up right now and asked to think about the sound of my son's name, that specific area—associated with language production and comprehension—would become activated. Just as with the previous point, this is not definitive proof, yet it once again raises the question as to where amodal symbols might reside. Because, if empirical evidence continually reveals well-established modal areas of the brain becoming activated, then there is at least equal reason to suppose the symbols have a distinctly modal structure rather than amodal.

The anti-amodal claim most supported by this study is the second claim: the grounding problem. If the traditional, amodal theory of cognition is correct, a subject in the Wheeler, Petersen, and Buckner experiment would receive sensory input to the visual system from the image and label, translate that information and concept into an amodal language of thought with modal semantic content, and store the resulting mental representation somewhere in the brain. Then, when asked to recall whether the word on the label matched a sound or image, the subject's cognitive processes would translate just the visible label into a language of thought and retrieve the mental representation (in Mentalese) associated with that formal translation. The retrieved representation would have encoded information that would indicate a visual or auditory concept without necessarily reproducing actual visual or auditory experiences, and the subject would then go through an output translation process to communicate his/her Metalese conclusions via the motor response of pushing a button.

The notion is similar to what occurs as music is encoded and decoded in digital files. The difference between the mental phenomena and the digital file is that we completely understand the translation process and mechanisms in the latter but not the former (if we presume cognition to be amodal). However, if the findings in the experiment are revealing a reactivation, simulation-like process in memory retrieval, the process and mechanisms are no longer shrouded in complete mystery. The stuff of cognition is based in, and composed of, reproducing some of the neural processes required to interact with the world (perception and action), and any translation is still grounded to modal activity and information.[15]

If the central processing unit responsible for memory retrieval is relying upon modal representations that trigger perception-like activity in the brain, then there is an explanation for how cognition interacts with perception and why mental representations retain sensory qualities similar to the initial stimuli. For example, memories of images seem like mental pictures because certain aspects of visual perception are initiated anytime one has a visual thought or memory. This does not rule out the need for translation processes within cognition. (Though it may seem like mental pictures, visual memories clearly are not actual images and the input must be encoded in some way to reside as a representation in one's body and brain.) But it does provide evidence to suppose the translation process and the resulting semantic content of the representations retain essential connections to the sensory systems from which they were produced. Thus making the representations themselves, and not just their semantic content, modal while also grounding those concepts—and resulting aspects of cognition which utilize them—in one's sensory systems and interaction with the external world.

Activation of Sensory Systems in Conceptual Processing

In addition to evidence of sensory systems becoming reactivated during memory retrieval of sensory-based concepts as described above, there is evidence showing similar occurrences when initially processing different concepts. This is a more contentious issue than memory

[15] This claim clearly seems to imply a form of empiricism, since the basis for all concepts comes from a subject's interaction with the world. The extent and implications of this empiricism are not essential to this discussion so it will be referenced only as needed.

retrieval because it leaves even less room for the potential of amodal symbols in mental representations. The previous experiment primarily revealed similar mental activity when processing and retrieving concepts. However, for supporters of the traditional model, such activity does not entail modality in mental representations, as it could be the case that the initial activities, and subsequent reactivations, are amodal translation processes. That is, the activity monitored in the fusiform gyrus when processing images, for example, could be the translation of information about external stimuli into an amodal, symbolic language of thought. Thus, seeing a reactivation of such areas—though still oddly dependent on very specific areas of the brain where the modal input is translated—need not be seen as conclusive evidence of inextricable modal properties in mental representations. If the initial neural activity is stripping the information of its modal dependency during the translation process, while retaining its modal semantic content, then any reactivity could be seen as amodal. But if the initial processing, translation, and encoding retains inherently modal information, physically and semantically, then the supporters of the traditional model will have a much more difficult task in arguing away findings such as those in the memory retrieval experiment. However, proving this is extremely difficult, especially when looking at abstract concepts.

If one is to accept the theory of modal mental representations—representations whose content and function always rely on sensorimotor processes—then at some point questions have to come up about amodal or abstract concepts. Concrete object concepts, such as APPLE, are easier to explain because they have properties related to sight, smell, taste, touch, and sound (like how a crisp red apple crunches). Abstract concepts, such as ARITHMETIC, will have only indirect relations, if any, to sensory systems. So, yes, maybe APPLE seems to rely on reactivations of sensorimotor neural processes, but we seem to have a great many concepts and propositional thoughts that have no overt ties to sights, sounds, smells, tastes, and textures.

Several experiments have attempted to uncover the neural source of abstract concepts by examining the location and dispersion of concept properties via fMRI. Two such experiments will be discussed; one looks at the neural effects of looking at food, while the other examines how subjects process the abstract concepts "CONVINCE" and "ARITHMETIC."

The food concept study was conducted by W. Kyle Simmons, Alex Martin, and Lawrence Barsalou (2005) and involved scanning subjects' brain activity as they viewed a series of images of places, food, and scrambled pictures. While viewing the images, subjects were instructed to indicate whether the current image was different from the preceding image via a response pad—the goal being that subjects could categorize what they were seeing effectively and thoroughly enough to have differentiation between the resulting concepts that would be mentally represented following each viewing. The results revealed that the subjects' gustatory cortex[16] became significantly more activated when they viewed images of food and did so consistently. Additionally, two regions of the left orbitofrontal cortex—believed to represent the reward values of tastes—became activated.

The significance of these findings is in the fact that the subjects were only shown images and not exposed to actual food. In the presence of real food such neurological activations could be non-conscious and non-conceptual, revealing only a reflexive physiological response to being exposed to something the body needs. However, given these were only images of food, the activations indicate a significantly modal characteristic to the concepts subjects represented. It appears that when one generates a concept—such as APPLE—it must include the property of

[16] The right insula/operculum. (Simmons, et al.: 1604)

being edible to the individual. This property is represented by activations in parts of the brain associated with motor-based, physical engagement with food. Had the researchers asked the participants to imagine eating the food shown to them perhaps this wouldn't be as striking. However, subjects were only asked to note differentiation between the images. So only minimal processing and categorization was necessary.[17] Thus, neural activity in the gustatory cortex and orbitofrontal cortex led the researchers to conclude that even minimal concept representation relies on various sensory and motor processing areas of the brain in order to effectively represent concept properties that indicate real or potential sensory and motor interaction. Thus, if a concept includes properties of edibility, the gustatory cortex will be activated; if it includes a graspable size, areas associated with motor control of the hand(s) will be activated; etc. As the researchers say,

> Consistent with previous findings, the experiment here indicated that conceptual representations are distributed across the brain areas that underlie their processing in perception and action. Because different categories are associated with different distributions of multimodal properties, different categories rely on different configurations of brain areas for conceptual representation…much work has shown that thinking about tools activates brain areas that process visual form, visual motion, and object manipulation. Analogously, we have shown here that thinking about food activates brain areas that process taste, taste reward and food shape. Thus our findings support the view that the brain areas representing knowledge for a particular category are those typically used to process its physical instances. (Simmons, et al., 2005: 1607)

Several others have looked at this experiment, and ones like it that cover other sensory modalities (Kiefer, 2005; Thompson-Schill, 2003; Gonzalez, et al., 2006), to reach the same conclusion. In a review of similar experiments and findings, Martin concluded, "the

[17] Though the researchers claim only minimal processing and categorization were necessary to complete the task, there is no way to observe or control a subject's involuntary mental actions. It could be that when a subject sees an apple s/he cannot help but imagine the taste or sound of biting into one. And if so, a supporter of the traditional model could says such sensual imaginings are linked to activations of motor and gustatory areas of the brain, and are being caused by the semantic content of edibility stored in an amodal symbol of APPLE.

neuroimaging findings…provide strong support for sensory-motor property-based models by revealing considerable overlap in the neural circuitry supporting perceiving, acting on, and knowing about objects…these findings suggest that object concepts are grounded in perception and action." (Martin, 2007: 27) By this "grounding," Martin is claiming object concepts aren't represented in a language of thought at all and instead emerge from activity within "property-based brain regions." And Barsalou has said, "When conceptual knowledge about objects is represented…brain areas that represent the shape and color of objects (fusiform gyrus), the motion they exhibit (middle and superior temporal lobe), and the actions that agents perform on them (premotor and parietal areas) become active to represent these properties conceptually." (Barsalou, 2008: 627) The consensus being that there is currently more than enough evidence to adopt a thoroughly modal model of mental representation when it comes to concepts of real-world objects. But what about abstract concepts?

As mentioned before, evidence regarding the modality of concrete object concepts is not nearly enough to prove all concepts rely on modal representations. And much of one's mental life is abstract. For one, language itself is abstract. Though it's possible to speak of syntax and semantics with great detail and as if the rules underlying them are universal, both involve relationships between symbols and meanings seemingly grounded in nothing more than relationships between symbols and meanings. (Such as how the definition of one word is established by one's understanding of the definitions of other words.) However, one could make the case that some words are grounded in concrete objects, when the words themselves describe, or are ascribed to, concrete objects. In this way, "red" can be communicated by one person pointing at a red object while saying "red" to another person. Later the listener may hear the sound of the word "red" associated with the letters R-E-D, and via inference connect the abstract language symbols to the concrete color of red. But many words do not designate concrete objects, and this is the challenge. Words such as "convince" and "arithmetic" do not have a one-to-one correspondence with any object(s) in the world. One way to explain how such abstract terms gain meaning is through amodal cognition: If cognition is really just a very complex form of symbol manipulation, then establishing the meaning of terms such as "convince" and "arithmetic" can be achieved via specific connections in the brain between certain sets of amodal symbols with the appropriate semantic content. It doesn't matter whether the concept is concrete or not. The properties associated with the symbols may be different for abstract versus concrete concepts, since the latter will involve a connection to sensory stimuli and possible motor actions, but, in the traditional model, sensorimotor input only affects the information that gets encoded and not the encoded symbol itself; the resulting representations, qua representations, are always amodal. At least philosophically, this is a very effective way at describing the myriad concepts human beings can fashion and possess. Thus, on the issue of abstract conceptualization, the burden of proof is on the defenders of embodied cognition to show how a modal model can account for the breadth of abstract thought and representation in the human mind. Unfortunately for embodied theorists, this is also the area most lacking convincing empirical evidence. But recent studies are beginning to show that modality may be instrumental in abstract concepts as well as concrete ones.

Perhaps the biggest challenge to establishing the modality of abstract concepts is finding effective means to examine the issue. Regarding concrete objects, the task is easier because it can be tied directly to a sensory system from the outset. Abstract concepts, on the other hand, seem to be tied only to other linguistic concepts (TRUST to LOYALTY; COUNTY to BORDER; etc.). Yet this very obstacle is what inspired one group of researchers to construct an experiment to prove the modality of abstract concepts: Wilson-Mendenhall, Simmons, Martin, and Barsalou. (2013) Their supposition was that if neural evidence could be found of abstract concepts tied to something other than linguistic content, then there might be evidence grounding abstract concepts in modal information.

> According to this view, abstract concepts are represented by situated conceptualizations that develop as the abstract concept is used to capture elements of a dynamic situation. For example, situated conceptualizations for the abstract concept *convince* develop to represent events in which one agent is interacting with another person in an effort to change their mental state. Any number of situated conceptualizations may develop to represent *convince* in different contexts (e.g., to *convince* one's spouse to rub one's feet, to convince another of one's political views). Thus, the lexical representation for "convince" is associated with much nonlinguistic semantic content that supports meaningful understanding of the concept, including the intentions, beliefs, internal states, affect, and actions of self and others that unfold in a spatio-temporal context. (Wilson-Mendenhall, et al., 2013: 921)

Essentially, the theory is that abstract concepts will be grounded in areas of the brain (and the process found therein) that retain content "relevant to social perception and interaction." (Wilson-Mendenhall, et al., 2013: 921).

In order to test this supposition, the researchers designed an experiment around four concepts, two of which were abstract (CONVINCE and ARITHMETIC) and two concrete (RED and ROLLING). The experiment required all subjects to undergo an fMRI scan while performing two different types of tasks. In the first task, subjects were shown scenes while being cued as to what to focus on in each scene. This included instructions to infer the thoughts of the individuals depicted in the scenes, count the number of independent entities in the scene, infer the direction of motion of an object in the scene, and infer the color of objects in black and white scenes. This portion of the experiment was designed to see if areas of the brain generally believed to be associated with the various cognitive tasks asked of the subjects (social cognition and emotion, number and magnitude, color awareness, motion detection) would become active in the subjects as expected. In the second task, subjects were shown one of the four concept words, asked to think deeply about its meaning for several seconds, and were then shown a series of scenes and asked to indicate via one of two buttons whether the concept applied to each scene. This process was repeated for all four concept words. After the initial task and scanning, several layers of analyses were conducted (functional localizer analyses, concept-scene matching task analysis, regions of interest (ROI) analyses, whole brain analyses, and split-half analyses) to ensure both accuracy and thoroughness in the measurements of neural activity.

The researchers predicted that subjects would show greater activity in brain areas associated with social cognition and emotion when processing the concept CONVINCE and greater activity in brain areas associated number and magnitude when processing the concept ARITHMETIC. The results proved the predictions correct. Great, but what does this mean? Though brain activity was not limited to only these areas when subjects were processing the abstract concepts, the fact that a higher amount of activity occurred in areas that have been established as centers of modal processing indicates that mental representations of these abstract concepts likely contain relevant nonlinguistic content. And, if even the slightest bit of modal information constitutes a given abstract concept, that concept cannot be said to be amodal. Thus, it appears these findings give embodied theorists reason to celebrate and traditional cognitive theorists reasons to pause. However, in this experiment it’s unclear if the modal connection is based on modal semantic content—which the traditional model allows—or on modal properties of the representation itself. Even the researchers who conducted the experiment are hesitant to pronounce these results as convincing evidence of a modal theory of cognition just yet. As they admit, “a challenge in studying abstract concepts is that the semantic content underlying their meanings is often quite complex…much remains to be learned about the variety of abstract concepts that people acquire and how they support thought and action in a complex world." (Wilson-Mendenhall, et al., 2013: 931) This admission is noted because their experiment was designed to test a very specific hypothesis regarding a small amount of semantic content on two abstract concepts. The findings are still quite significant for the claim that all cognition *might* be grounded in some way with sensorimotor systems and neural processes that accompany them, but it is still uncertain if this can be applied to all abstract concepts. Yet it once again reveals a lack of evidence of modal semantic content being translated into a wholly amodal language of thought, as the centers of activity while performing the tasks were well-established modal areas of the brain.

The separation of color knowledge from color perception

This last heading may sound as if it is related to an argument in favor of amodal cognition, but the findings discussed in what follows help further the claim that some semantic content must be grounded in and continually coupled with modal systems. Miceli et al. (2001) discovered that one can possess accurate color perception and naming abilities, but if those cognitive abilities are unable to communicate with other parts of the brain due to brain damage, one may lose the ability to associate certain colors with certain objects, such as yellow with lemons and orange with carrots (what the researchers refer to as "object color knowledge").

The experiments conducted by Miceli, et al., involved two subjects who suffered from a very unique form of achromatopsia due to damage to the mesial temporal regions of the left hemisphere, wherein they seemed to possess the ability to accurately name colors displayed in front of them, yet could not apply colors accurately to certain concepts and objects. The former was fairly easy to establish by showing the subjects colored plates and asking them to either name the colors or arrange the plates to match with specified colors. The subjects were able to do this successfully. This established that both subjects were able to perceive colors effectively. The second part of the experiment involved several exercises where the subjects were asked to answer questions regarding colors associated with certain concepts (such as, "Can a lion be red?"), asked to pick the canonical color associated with objects (such as lemons and carrots) from a selection of crayons, and shown black and white images of objects such as a chair, bicycle, pumpkin, and others, and asked to both name the object and the canonical color (if any) associated with it. The results of the second portion of the experiment showed significant impairment in the subjects' abilities to correctly apply colors to the objects presented to them. That is, the subjects correctly identified the pumpkins as pumpkins, but did not associate the color orange with that object. Similarly, though they could identify a yellow swatch of color as yellow, they were not able to associate yellow as the canonical color associated with lemons.

The researchers believe these findings reveal a unique dissociation between active perceptual knowledge and stored conceptual knowledge. This dissociation was caused by the subjects' inability to link the two processes (due to brain damage). Without the ability to access or utilize the neural mechanisms involved in perceiving color, the areas of the brain associated with stored properties of objects could not effectively apply color properties to concepts—possibly indicating the need for conceptual properties to be grounded in modal systems, such as vision, in order to be applied effectively in cognition.

> The observation that access to object color knowledge can be damaged independently of access to object form knowledge is consistent with those theories of the organization of conceptual knowledge that assume that such knowledge is distributed across various 'modality'-specific systems, such as visual, auditory, and motor, functional and verbal systems…the data we report here suggest that semantic information about the perceptual properties of objects must be grounded in modality-specific neural systems. (Miceli, et al., 2001: 666)

If correct, this could mean one's knowledge of concepts with modal-related semantic properties—such as color, texture, smell, and sound—could be impaired if the storage center of the concept cannot communicate with the necessary sensory system. One might lack the association of red with stop signs, the property of roughness with sandpaper, sweetness with honey, etc. They would still possess concepts regarding objects like stop signs, complete with its function, shape, placement near intersections, and so forth, but would lack (sometimes essential) modal properties of the concepts.

Of all the experiments discussed so far, this is perhaps the weakest at present. For one, there isn't yet complementary evidence that shows similar dissociations between perception and knowledge of all other modal concepts, such as the roughness of paper or the sweetness of honey. Color is the only modal property shown to be affected. But that may be due to the lack of similar cases; the brain damage of the subjects is extremely rare. However, that does not undermine the possible implications this experiment reveals on modal semantic content's ties to modal sensory systems. But this leads to another weakness: since the problem is in modal semantic content of already established concepts this could be an issue of knowledge impairment rather than concept impairment. The subjects in the experiment showed an awareness of the concepts PUMPKIN, LEMON, CARROT, CHAIR, etc., as well as an ability to recognize various colors concepts. The problem was in knowing what colors were connected with certain objects—the problem of associating one perceptual concept with another object concept. The ability to fully flesh out a given object concept with all modal properties is obviously something that could be beneficial to a person, but the inability to do so in the study may only show how cognition is coupled to sensory systems in one correlative aspect of cognition (the ability to correlate color concepts with object concepts) rather than cognition as a whole, since the subjects were still able to retain both object and color concepts. But, even in the face of these concerns, the value of this experiment is more in its how it adds to the continuously-growing list of findings indicating links in cognition that undercut the traditional model yet are explainable in an embodied approach. The hope of the embodiment theorists is that the more links found tying cognitive functions to sensory systems, the closer we get to recognizing the entire structure upheld by mind-body-world interrelations. So that with something like a pumpkin, perhaps it will eventually be shown that all essential non-color properties that also make up the concept PUMPKIN—its size, texture, shape, taste, smell, expectation of location, etc.—will be shown impossible to relate to an object (conceptually and perceptually) without specific sensory systems actively involved. The trouble will be trying to achieve this without asking subjects to relate modal properties to already established object concepts because such a procedure inherently relies on the notion that one can possess and access concepts independent of whichever sensory system is involved in the modal property being examined.

Really, the entire argument that representations are not amodal carries this same benefit and concern. None of the evidence is definitive, it complicates the traditional model but doesn't necessary rule it out, and some issues (such as the modality of abstract concepts) are terribly difficult to test at all. Yet, the findings of these experiments and ones like them consistently align with predictions offered by embodied (or at least non-traditional) theories. In that sense, they are fuel for further discussion on the issue. Additionally, the claims mentioned at the beginning of this section continue to be legitimate concerns against amodal cognition. There is an ongoing lack of clear evidence of amodal processing, no fulfilling explanation for how an amodal central processing unit interacts with the sensory and motor systems, and a lingering question as to where amodal processing occurs as nearly all cognitive tasks observed showed a definite overlap with modal processing areas.

For better or worse, there is much more that could be discussed on this topic. Does it really undercut the traditional model? Is the relationship between sensory systems and modal semantic content metaphysically or nomologically related? Does the empirical evidence support anything more than a thoroughgoing empiricism? Some of these questions will be addressed in more detail in chapter 4 when discussing the main problems plaguing embodied accounts of cognition. But for now, we'll move on. In the next section, on language, we will see why some theorists have philosophical and empirical reasons to suspect all linguistic symbols, both abstract and concrete, are the result of perceptual systems.

3.4 Language is based on embodiment

Initially, language may seem like a wonderful analogy for traditional cognitive science. You have symbols designated to represent everything imaginable—concrete or abstract—with each symbol or grouping of symbols having meanings and roles determined by the underlying rules of the system. Plus, written and verbal language is arbitrarily symbolic. In different languages, similar shapes connote different sounds and similar sounds are depicted by different shapes—none of which is based on any universal standard of available shapes, sounds, methods, or even reading direction. Symbols, sounds, and words have the meanings they have based seemingly on nothing other than the agreement of those people who are using the language. This arbitrariness is important because it frees the language from being grounded by or limited to one's embodiment. Other than the requirement of being seen, heard, or both (and usually required to be public), language is a symbolic translation process that can operate effectively regardless of the particular physical form of the users of that language. Nevertheless, some have looked at language and found what they believe to be evidence for embodiment.

In this section I will look closely at those who believe language is one of the most perplexing issues for the traditional model and the most compelling evidence for embodiment. Perplexing because, similar to the grounding problem discussed in the last section with mental representations, there needs to be a non-circular explanation as to how language symbols come to have the meaning they have. The compelling evidence is that a good amount of language comprehension seems reliant upon motor processing systems within the brain as well as metaphors based on embodied concepts. In regards to the former, this section will discuss the possibility that language may not be innately, arbitrarily structured and is instead reliant upon embodiment, and the sensorimotor affordances embodiment entails, for it to be effectively communicated and understood. In support of this, I will introduce Arthur Glenberg's indexical hypothesis, which claims linguistic symbols become meaningful through a process in which words are mapped to "perceptual symbols" based on affordances derived from those perceptual symbols. Additionally, several neural studies will be discussed that seem to reveal a significant relationship between motor systems and language comprehension and usage. In regards to the role of embodied metaphor in language, theories will be presented which claim non-neural structures foster, determine, and possibly constitute the acquisition and use of language. The most prominent figures covered will be George Lakoff and Mark Johnson and their argument that the mind is biological rather than symbolic, and we can recognize the essential nature of embodiment in thought by examining the structure, use, and role of body-based metaphors utilized in language.

The Source of Meaning

John Searle's famous Chinese Room thought experiment nicely captures the grounding problem facing language. In his hypothetical scenario, we are asked to imagine a man inside a small room, containing nothing other than a rulebook for manipulating Chinese language symbols. Each side of the room also has a slot. On one side, messages in Chinese are sent into the box. He then consults the rulebook in order to determine the appropriate response to the messages (also in Chinese), which he then transcribes and sends out the slot on the opposite side of the room. The crux of this scenario is that the man in the box has no understanding of the Chinese language. He doesn't know how to speak, write, or read it, and therefore has no idea what the content of the messages he is receiving or sending are communicating. Yet, at least from the outside, it could appear as if the man in the box understands Chinese because he is receiving messages and replying with appropriate responses. He seems to be utilizing and processing the language adequately. However, the language has no grounding. The symbols don't actually symbolize anything for him; they are simply distinct shapes that require a response of other distinct shapes based on the instructions in the rulebook. The concepts, objects, emotions, people, times, places, etc., described or referred to in the messages are not recognized or even accessible to the man in the room. Without some connection, or grounding, between the symbols and what they are intended to symbolize, the language is meaningless for him.

On the surface, this basic problem applies to computationalist cognitive models as well. Though the symbols in Mentalese may seem to work together effectively, if the meaning of the symbols is not grounded in some way to what the symbols represent, the results are meaningless to the mind housing these mental processes. Taken on its own, this problem is not an argument in favor of embodiment; it merely shows that computationalism and traditional cognitive science have some questions to answer in regard to where meaning or understanding comes from. This only becomes a point in favor of embodied cognition if embodied theorists can provide a successful account of the source of meaning. According to Glenberg's indexical hypothesis, it just might.

But before jumping into a discussion on Glenberg and the rest of the language debate, here's a note of warning: "meaning" is ambiguous and the philosophers involved in the traditional account versus embodiment debate are not all using "meaning" in the same way. This disparity has great significance since this particular aspect of the embodiment argument is generally focused on foundational theories of meaning. That is, they are trying to determine what basic relationship or process gives symbols the meanings the have. However, several of the traditional accounts embodiment takes aim at are not. Instead, many of them are concerned with semantic theories of meaning, looking at the meanings mental symbols have and how they operate with other mental symbols. Both approaches are interrelated but each is seeking to answer fundamentally different questions. Consider the Chinese room thought experiment—when asking how the symbols acquire meaning for the individual in the room, there is an explanatory gap. Proper function alone does not account for intentionality. But if one were attempting to discern how the symbols operated—the rules dictating connections between combinations of symbolic input and output—his/her questions could be answered via the room and rulebook alone. The embodiment theories about to be discussed argue that a semantic evaluation of meaning is not enough. They say the meanings, inferences, and conceptualizations derived from language must have a source outside of the language itself. Otherwise, intentionality cannot be explained. This is highly debatable—both whether intentionality cannot be explained without foundational grounding and if intentionality is essential to establishing meaning within a language. Unfortunately, the debate between foundational and semantic theories of meaning is far too much to cover in this work. But it is worthwhile to keep this approach and bias in mind from the start.

Glenberg's Indexical Hypothesis

In short, Glenberg's hypothesis claims linguistic symbols become meaningful by "meshing" perceptual symbols with sensorimotor affordances via a three-stage process:

Stage 1: Words and phrases are indexed (or mapped) to perceptual symbols.

Stage 2: Affordances are derived from the perceptual symbols.

Stage 3: Affordances are meshed with linguistic symbols, thus yielding an understanding of the word or phrase.

The notable difference here, as compared to traditional amodal theories, is that perceptual symbols become the crucial element as opposed to amodal symbols or concepts. Fodor has said that in the language of thought, "just as the semantics of sentences are constructs out of the semantics of words, so the semantics of thoughts are constructions out of the semantics of the concepts that are their constituents." (Fodor, 2008: 19) Since Fodor takes an atomistic view of semantics, he is saying that concepts themselves—as mental objects/representations with mental particulars not derived from definitions or inferences—provide "meanings" for words or phrases. And as these mental objects do not reside in a perceptual system or require the reactivation of a perceptual system to be utilized in cognition, the overall theory is considered amodal. Glenberg's use of *perceptual* symbols is then supposed to clearly indicate that words and phrases are not mapped to an amodal system of thought, but rather a thoroughly modal one. But this phrase is vague, so more must be explained about perceptual symbols in order to understand the impetus and impact of Glenberg's hypothesis.

The term "perceptual symbol" comes from Barsalou (1999) but was adopted by Glenberg as well. Barsalou claims perceptual symbols are modal because they are not represented in a system or area of the brain independent of the perceptual system(s) responsible for their creation. The activation in perceptual systems when accessing or utilizing representations of concrete objects and other sensory concepts can be interpreted as a sort of simulation of the original perception, thus making the semantic content of the representation analogous to act of perception. With this in mind, we can begin to look at and understand the three stages of the indexical hypothesis. The first stage is fairly straightforward: since the words (both in sound and symbol) ascribed to a given object are arbitrary and not inherent or universal, the object must be christened in the language of the perceiver and designated by a specific word or phrase. Thus, one might see a chair for the first time and another speaker might point at the object and say, "chair," with the goal being the first-timer perceiver will generate a mental representation of the object that includes the sounds and symbols that make up the word "chair." In the second stage, as the perceptual symbols are formed, one derives the ways in which one can physically interact with the perceived object. In the case of a chair, one might recognize that it offers a place to sit, raised off the ground. If it has arm rests, one might see them as something to grasp or rest forearms upon. If it has a back, one might recognize it as a place to both sit and lean back, or even an object on which to drape other objects. If it appears to be sturdy, one might see it as a stool-like object to stand upon in order to change a light bulb. If flimsy or able to rotate, it may be perceived as a place where one should *not* stand, etc. As the affordances are recognized and categorized they are meshed[18] with the linguistic symbol(s) in the mental representation, imbuing the word or phrase with sensorimotor possibilities and limitations. The resulting linguistic symbol has the meaning it has based on one's potential interactions with it, not just on the abstract rules governing the language. This is why Glenberg says, "to a particular person, the meaning of an object, event or sentence is what that person can do with the object, event, or sentence." (Glenberg, 1997: 3)

The affordances and the meshing that occurs then give meaning to sentences that would otherwise be nonsensical, even if all rules of grammar and syntax were satisfied. Consider the following examples:

1. Because she was cold, Chloe put on her coat.

[18] The "meshing" process is unfortunately unclear. Somehow this is the point where the linguistic symbol and affordances become coupled. And though the affordances can exist mentally without having a linguistic symbol, it is described as if the linguistic symbols cannot exist decoupled from the perceptual symbols. Or, if so, they cannot be considered "linguistic" symbols. They only work in language based on the relationship they have with information tied to them from perceptual symbols.

2. Because she was cold, Chloe put on her fish.

Both sentences are grammatically correct and provide the reader with understandable notions of Chloe's actions. However, the second sentence doesn't make sense because fish are not associated with warming one's body. Because "coat" has been meshed with affordances related to warming one's body, the sentence makes sense when Chloe puts on her coat yet seems absurd when she puts on fish due to the cold—due to fish lacking any affordances related to body temperature. Similarly, if one did not know the affordances associated with a given term, the sentence would be incomprehensible. If, for example, instead of "coat" the noun was "Mantel," and the reader did not know "Mantel" was the German word for "coat," the sentence might be meaningless. And if it weren't—if the reader deduced that "Mantel" might mean some equivalent of coat—this would likely be due to inferences derived from the reader's understanding of objects one might use to respond to cold, based on the specific affordances of those objects.

Additionally, this is thought to be somewhat different to Frege-like scenarios where a speaker's sense of coat does not align with his/her sense of Mantel, even though they refer to the same thing. Problems such as this can occur due to a misunderstanding or lack of awareness of the definitions of the terms. If that were the case, the fault would be in the properties one associates with concepts linguistically. For example, the concept COAT might encompass properties of 'ability to warm or keep warm,' 'ability to cover upper body,' 'ability to cover arms,' etc., and one could survey the concept and make inferences that would determine if "coat" is suitable to the scenario described in the sentence.[19] Thus, whether "coat" or "fish" are

[19] It should be noted here that Fodor does not find this 'definition-in-use' approach as acceptable in the language of thought either. He explicitly rules it out, in part because it never explains what

appropriate depends of the definitions of each term. Instead, for Glenberg and Barsalou, the inferences are the result of affordances derived from perceptual symbols and not from linguistic definitions. The difference is that "perceptual symbols capture the projectable properties that are relevant for the actions we are contemplating, and in that sense perceptual symbols capture relevant affordances." (Glenberg, 1997: 47) It is an opposite direction of fit than a definitional explanation of inferences. According to the indexical hypothesis, language isn't mapped onto the world (where one starts with the definition and then finds its referent); the sense one has of a given word is based on the experience of the referent and the non-conscious extraction of affordances via the various sensorimotor systems. As Glenberg states,

> Importantly, embodied representations do not need to be mapped onto the world to become meaningful because they arise from the world. In other words, embodied representations are directly grounded by virtue of being lawfully and analogically related to the properties of the world and how those properties are transduced by perceptual-action systems. (1997: 3)

In this way, the meshing that occurs allows users of a language to discriminate between appropriate and inappropriate uses of terms and phrases within the language based largely on physical potentials derived from experience, and not just definitional and inferential rules applied to experience.

If acceptable, this hypothesis has several compelling implications for linguistic symbols. Since the affordances are derived from the perceptual symbols, which, in turn, are continually dependent on the processes underlying the original perception(s), language is understood based on re-activating or re-accessing the original, modal perceptual system(s). The meanings of words are then not bound to descriptions that rely on arbitrary systems utilizing amodal symbols. If they

---

holds sense and reference together. Where Glenberg is utilizing meshing and affordances to do so, Fodor says such neo-empiricist approaches never provide a non-circular explanation of how the rules governing inference in concepts arise.

were, there would be no reason why activations in one area of the brain would be necessary to represent a given concept rather than another area of the brain. According to Glenberg (and Barsalou), the concepts are bound to specific neural responses and processes of sensory systems. This is the so-called "analogical" nature of perceptual symbols and it requires that they reproduce or simulate an experience like that of the original perception. They believe this cannot be achieved in an amodal mental medium. "We understand language by creating embodied conceptualizations of situations the language is describing." (Glenberg, 1997: 12) By "embodied" conceptualizations, Glenberg intends concepts based on experience of what one has, is, or could do. This goes back to his claim that the meaning of an object, event, or sentence is what a person can do with that object, event, or sentence.[20] Concepts are inextricably bound to experience to the point that thinking of or even non-consciously utilizing a given concept entails re-experiencing it. And since a fully amodal concept is supposed to be accessible without necessarily relying on reactivating analogical neural processes, the amodal model must be abandoned.

Furthermore, affordance-based concepts allow one to recognize potential action associated with objects for novel situations as well as existing ones. This can be understood by looking a term like "ladder." A ladder can be defined as a structure of wood, metal, or rope, commonly consisting of two sidepieces between which a series of bars or rungs are set at suitable distances, forming a means of climbing up or down. The first part of the definition is material, but the last part is an affordance. If the affordance becomes the critical point, then ladders can be

[20] And this goes back further to the distinction between versions of "meaning" mentioned at the beginning of this section. When he uses the phrase "what a person can do" he's not describing how one can compose propositional attitudes or concepts based on possible relationships between mental representations. He is describing physical interaction with the world and claiming the source and purpose of meaning always comes from and often returns to this interaction.

found in a great many objects, some of which may not include the common material components in the first part of the definition. Therefore, one might encounter a pile of rocks and if the pile offers the potential motor interactions of climbing up or down then one might recognize the pile of rocks as ladder-like. This is due to properties derived from perceptual symbols and the affordances therein—the recognition of properties such as can-climb-up-or-down is not limited to objects physically similar to common ladders, but to objects that offer a similar affordance. This allows objects and their potential functions to be determined by the sensorimotor affordances they offer rather than the descriptive traits assigned to them through linguistic symbols alone. This is why, in the absence of a ladder, one will utilize disparate objects to achieve the same end of climbing up or down (such as a table, flipped-over pot, stack of books, etc.). One begins to recognize a fit between properties of the objects and embodied knowledge. Even though those objects, by their linguistic definitions, are not objects with the property of a means to climb up or down, the affordances associated with their shapes, sturdiness, and height allow one to use them in novel ladder-like ways.

This is also what allows one to use and understand verbs and adjectives in sentences that offer no clear opportunity for action. Many sentences one can understand have no obvious relation to what one (or any other human being) can do with that sentence, object, or event. For example, "The quick brown fox jumped over the lazy dog" does not directly communicate any possibility for action in the world for the writer or reader. However, similar to the extrapolation of ladder-like affordances found in non-ladder objects, Glenberg says one can understand sentences like the one above for two reasons: 1) We interpret human-like affordances in non-human objects, and 2) We conceptualize *all* concepts based on perceptual symbols. As Glenberg says, "it is our embodied interpretation of the situation that forces (or at least makes probable) an interpretation of the subsequent language." (Glenberg, 1997: 48)

This takes many forms in this particular sentence. For one, the representations of "fox" and "dog" carry with them certain sensorimotor affordances, even though they are not as obvious as in sentences about coats and fish. Though FOX and DOG as working concepts may eventually develop myriad criteria for what separates one concept from the other phylogenetically, how one can interact or has interacted with the actual animals will partially determine how one conceptualizes them. For example, when approached, both foxes and dogs are expected to be smaller than an average adult human. So, when reading the sentence, an expectation of size arises based on one's embodied experience of relative height differences.[21] Likewise, the brown fur will be conceptualized as having a certain texture based on how one has previously touched or seen animal fur or fur-like materials. "Quickness" is a trait understood based on one's movement in the world or the observation of another object going from point A to B in a comparatively brief amount of time. Similarly, "laziness" is understood based on inactivity. However, it does not follow that "laziness" or "quickness" has a specified set of characteristics similar to a definition.

These terms can be applied to objects or non-human animals based on our embodied understanding, but they are *interpreted* within the situation described in the sentence. Thus, one can read and understand the sentence without ever having observed a brown fox moving quickly or a dog behaving lazily; the sentence is comprehended based on the manner in which it is

[21] Again, this could be based on experiences with actual foxes or depicted foxes. If one has seen an actual fox then it has been experienced embedded in an environment within which other objects seen in relation to it can provide a sense of scale. Additionally, if one had only seen images of foxes in storybooks, scale will be provided within the images. If for some reason the storybook depicted the fox as much larger than a human being then one would presumably conceptualize the fox in the sentence above as unusually large.

conceptualized. And the conceptualization is derived from embodied experiences and the projection of those embodied experiences onto non-human animals. Glenberg predicts that if we were to continue examining the understanding of this sentence and unpacking the terms, we would continue to uncover either affordances or perceptual information derived from embodied experience.

Additionally, our conceptualization of the sentence can evolve over time. As one perceives foxes in action or dogs lying about, the conceptualization may change based directly on perceptual symbols. But until that occurs, the sentence describing the quick fox and lazy dog (and others like it) will be understood by meshing the information from stored perceptual symbols with the words and scenarios of this particular sentence. This, Glenberg believes, shows that we continuously rely on embodied interpretations to comprehend or establish representations that are, in his words, not fully embodied.

These implications are significant, yet one vital aspect has not been discussed yet: the meshing process itself. How do the affordances, once derived, combine with the linguistic symbols to give words and phrases meaning? First off, one will have to grant the possibility of embodied conceptualization, meaning, and memory. In addition to what has been stated earlier, Glenberg's rationale also aligns with claims and arguments set forth in the sections on perception as perceptually-guided action and modal representations. He echoes Gibson, O'Regan, Noë, and Hurley when saying, "meaning of an object or a situation is a pattern of possible action…determined by the projectable features of the object molded by bodily constraints and modified by memory of previous actions." (Glenberg, 1997: 4) Thus, he views meaning to be shaped and guided by embodied interaction with one's environment and implies that relationships among amodal symbols is not enough to provide meaning. Just as in the Chinese room, symbol-to-symbol relationships have operational value but lack "meaning" until there is some sort of intentionality grasped by the person in the room. Glenberg seems to be reiterating Searle's claims that meaning requires intentionality, with the added requirement that intentionality requires embodied interaction with the world. This interaction then "grounds" meaning by providing both a source and direction for all conceptualizations; all meanings are derived from and continually dependent upon perceptual symbols so meanings will be shaped on how one can (and wants to) interact with the world.

To explain this, Glenberg uses the example of encountering a path in the woods while heading home. When attempting to get home, one will look for certain patterns of potential action such as direction, the ability to fit one's body between shrubs, the height of branches impeding progress, the slickness of the surface, not going through creeks if one wants to remain dry, etc. These goals, accompanied by one's background knowledge about one's height, ability to traverse certain obstacles, and so forth (stored in perceptual symbols), as well as one's ambition to get home, mesh with the "path" being perceived in order to consider it a homeward path. If the supposed path lacks these affordances, one will not consider it a path and will look for one that can be successfully meshed with these patterns of interaction.

> In conceptualizing the environment as a path, the spatial-functional patterns based on projectable properties from the environment are combined or meshed with the patterns from memory. The meshed pattern dictates how (or if) the body can be moved in a way that simultaneously satisfies both sets of patterns of action (e.g., "Can I, with my body, get from rock to rock without getting wet?"). This sort of mesh is a possibility because all of the patterns are embodied, that is, they are all encoded in terms of how your body constrains actions. When the patterns can be meshed into a plan for coherent action (e.g., stepping across the rocks), the rocks, soil, and twigs become (for you) a path. (Glenberg, 1997: 6)

This is a helpful explanation but the process of meshing is still mysterious in terms of the neural processes underlying the cognitive activity described by Glenberg. However, Glenberg and

Michael Kaschak have conducted several experiments that support this hypothesis in terms of sentence comprehension and its grounding in action.

In order to establish evidence of this notion that "language is made meaningful by cognitively simulating the [embodied] actions implied by sentences," Glenberg and Kaschak conducted experiments designed to show how the action(s) one infers from a sentence can interfere with an action one is attempting to complete. (Glenberg and Kaschak, 2002: 559) The hope was that these uniquely constructed experiments could reveal an embodied influence of sorts on various cognitive tasks. The key to the hypothesis resides in the neural simulations that accompany accessing information tied to perceptual symbols. If a subject reads a sentence describing an activity involving, say, the left hand then it is believed the neural processes for left-handed movement become activated. So, in a way, the subject is readied for action with the left hand as a result of reading and understanding the sentence. Relying on this notion, the experimenters wanted to see if subjects, who were given response pads that would be used by hand, would have an easier time responding when the actions described to them would match (somewhat) the actions required to provide the correct answer on the response pad. For example, subjects were required to respond with the left hand to sentences describing left-handed actions, versus requiring them to respond with the right hand to sentences describing left-handed sentences. If subjects showed significantly quicker responses consistently in instances where response actions could be done with the hand described in the sentences, then the indexical hypothesis could explain such findings by claiming the activations of neural processes for simulations (for conceptualization) would "prime" a subject and allow him/her to complete the physical action quicker than those whose simulations did not match the required physical action.

This was very similar to the experiments discussed in the last section. The difference here was an emphasis on processing language rather than perceptions.

In one experiment, subjects were given several sentences and asked to determine if they were sensible or nonsense, and then push a button indicating their response. The sentences were never as direct as "Push the button farthest from you," or "Use your left hand to touch the nearest button." Instead, the action was described indirectly via various directives. Sentences such as "Open the drawer" and "Put your finger under your nose," indicated action toward the body; whereas sentences such as "Close the drawer" and "Put your finger under the faucet" indicated direction away from the body. The nonsense sentences did not indicate any direction but included commands such as "Boil the air." The implied directions were important because subjects had to respond on a specially constructed button-box with three buttons perpendicular to the subjects' bodies, which put one button farther away from the subjects' bodies, one much closer to the body, and one in between. Subjects were required to press the middle button until they were ready to view the sentence. Once the sentence was visible and read, they would press one of the near or far buttons to indicate whether the sentence was sensible. Additionally, as one of the variables of the experiment, the placement of the "Yes" button would change. In some instances, the "Yes" button was the farthest away from the subjects' bodies, in others it was the nearest. The researchers measured the amount of time it took subjects to read the sentence and select the appropriate response.

If the indexical hypothesis is accurate, then meaning is grounded in action. Thus, as mentioned before, Glenberg and Kaschak hypothesized that in order to understand a sentence such as "Open the drawer," which requires pulling a hand toward one's body, one simulates action towards the body in the neural states necessary for completing such an action. (Likewise for actions pushing hands away from one's body implied by sentences such as "Close the drawer.") And if these cognitive simulations require activation of the same neural systems for accomplishing real action, the simulations could interfere with the real action if it is not aligned with the sentence. The idea being that subjects would react slower when the sentence indicated away action and pushing the "Yes" button required coming toward the body, and vice versa. They called this interaction between implied sentence direction and actual response direction the action-sentence compatibility effect (ACE).

The results showed ACE as having a measurable impact. When real action aligned with implied sentence action—when "Yes" was away from the body and the action implied in the sentence was too, or both were near—the response time for near-near was generally faster than far-far. However, when the button and the implied direction were opposing, both near and far responses showed an increase in response times of up to 100 milliseconds. The response times varied depending on the types of sentences shown (some involved only the subject and a physical object, while others involved multiple subjects, and still others involved nonphysical nouns) yet all showed the same effect. Glenberg and Kaschak believe this to be strong evidence that some language understanding relies on action-based neural systems. And, importantly, this applies to concrete as well as abstract notions. Though some sentences referred to physical interaction with drawers, sinks, and other objects, other sentences referenced abstract actions with things like messages and stories ("Liz told you the story," "You radioed the message to the policeman"). This is possibly the most significant finding in the experiment because it helps address how even abstract notions—which would seem to be much more difficult to explain in an action-based theory, and perhaps less so in a wholly abstract, symbol manipulation theory—require grounding in bodily action. Though the traditional model may have a response to the interference when it correlates to sentences involving motor responses with concrete objects (discussed below), the traditional model cannot make sense of this occurrence with regards to sentences involving abstract nouns.

In regards to sentences involving concrete objects and motor responses, one could attempt to explain the interference in terms of cognitive priming. Perhaps, due to background knowledge associated with the verbs and directions communicated in the sentences, one's body is put on alert, so to speak, when exposed to those words. So that when one reads the word "Jump," for example, in an imperative sentence that demands action—such as "Jump off the step"—the brain begins to activate processes simulating the action in order to facilitate the commanded action. A subject may choose not to jump off the step, but the neural processes necessary to accomplish the task may still have been mildly activated due to a sort of preparation for action rather than due to a reliance on action processes in order to understand the command. And once primed, the redirection necessary to push a button when it is opposed to the direction implied in the sentence will cause a slight delay. It is possible that such an occurrence could be explained by the traditional notion of amodal symbol manipulation. They could say the output, or preparation for output, may involve motor processes, but the cognitive processes responsible for priming, redirecting, and triggering them could still be purely amodal.

However, this explanation cannot apply to abstract sentences in which there is no command or description of actions that might need motor priming. The verbs "radioing," "messaging," or "listening" do not require one to actually move toward or away from one's body, there is no reason to suppose one's actual motor responses would vary as they did with the concrete sentences. Yet the interference was present when subjects responded to concrete *and* abstract sentences (it was more pronounced with concrete sentences, but consistently present with both). Interference should only happen when a primed motor response opposes the necessary motor response, causing a slight delay as the motor response is redirected. One doesn't really need to move toward someone to send a message and one doesn't need to bring anything toward oneself in order to hear a story. This is inexplicable under the traditional model, but it is consistent with the indexical hypothesis and its claim that "real bodily action is at the root of meaning conveyed by language" (Glenberg and Kaschak, 2002: 563) And not just some language, but all language. This is easier to accept for concrete verbs and objects that refer to sensorimotor input and output, but more explanation is needed to show how and if all abstract notions fit in.

The experiment described above (as well as others performed by Glenberg and Kaschak) addresses only a handful of abstract sentences, all of which are not too far removed from literal action. This leads one to ask why Glenberg and Kaschak are so confident of the ubiquity of the indexical hypothesis given the massive scope of possible abstract terms in language, many of which seem to be much further removed from action than those used in the experiment. Though a full-fledged empirical analysis is not easily attainable (or maybe possible at all), they believe abstract concepts and actions can still be explained under the indexical hypothesis by looking at a few paradigm examples.

First, consider a sentence like "That is a beautiful sunset." The traditional model could say this sentence has meaning based on the definitions of the words and the relationship between the noun and adjective based on the syntax of the language. A sunset relates to the position of the sun at a certain time of day and beauty is an abstract qualifier of a given object indicating that it is extremely pleasing in some way. On the surface, nothing about this sentence seems to be tied to real bodily action. But Glenberg and Kaschak believe this sentence can be understood as providing a new body-based perspective within a rich context. That is, it communicates a new affordance to a subject about an experience already understood in real-world contexts. Though the affordance is quite subtle here, when one understands the sentence "That is a beautiful sunset" it relies upon one's background knowledge of actually experiencing sunsets and calls attention to the fact that this sunset affords one the opportunity to look at it and experience some form of aesthetic pleasure. "The sunset affords looking at, and acting on this affordance results in the goal of a pleasurable experience." (Glenberg and Kaschak, 2002: 563) In this example, there is an assumption that one has had a pleasurable visual experience before and can then have the goal of experiencing this pleasure again, which will affect the affordances one associates with the scenario described in the sentence and thus the meaning of the sentence to the reader or hearer of the sentence. If beauty (understood as a sort of pleasurable experience) cannot be associated with the noun because it does not afford the opportunity for a pleasurable experience then the sentence will be nonsensical. For example, "That is a beautiful noun" could be a vacuous statement to many[22] because an unspecified noun offers no indication of affordances. But the sunset, associated with sight (and possibly other sensations) has affordances tied to perceptual symbols.

Another example involves the even more abstract notion of causal reasoning. In terms of pure observation, as Hume famously pointed out, one only observes events immediately followed by other events. Thus, linking one event to another in a causal relationship is a matter of inference. But when one makes a statement such as "Jackson pushed the pedal, making the car accelerate," or "The heat from the stovetop is causing the water to boil" it is difficult to make sense of the causal relationship depicted if the language of causation is grounded in nothing other

[22] I'm leaving the door open a little here because it could be that passionate grammarians can experience pleasure from noun alone.

than amodal symbols. It is *difficult*, though not impossible. One could derive an inference of causality from an analysis of the covariation of events (Novick and Cheng, 2004), but Glenberg and Kaschak believe an embodied analysis is much simpler. Similar to the claims of O'Regan and Noë about how infants learn how to adjust their bodies in order to bring about new views and sensorimotor possibilities by moving body parts with sufficient force to affect other objects, "infants learn to conceptualize causes as the application of a bodily force to effect a change or resist a change." (Glenberg and Kaschak, 2002: 564) This experience and the concepts resulting from it then ground all further notions of causation to the point that even as an adult one uses and understands causal language (for physical, social, and psychological causes) as various pushes and pulls derived from one's bodily experience. This is why storm fronts are described as *pushing* across the landscape, why dry rice is thought to *pull* moisture from a waterlogged phone, and why sentences communicating disgust imply (and as they will say, rely on) repulsion, and sentences communicating desire imply attraction.

This is quite a leap, having gone from affordances related to verbs directly describing one's action in the world to affordances related to understanding abstract first-personal verbs, now to affordances and sensorimotor activity providing the basis for conceptualizing causal relationships between natural kinds. An important thing to keep in mind here is that they are not yet claiming causes to be real events. "Cause" is simply a word that is being mapped to a perceptual symbol or symbols. Going back to Hume, one may only perceive successions of events without ever directly witnessing a cause. For Glenberg and Kaschak, causation is a notion and term ascribed to certain successions of events. When a child pushes or pulls on something, causation is not part of their perceptual information. But a chronological order of events is. A child extends a hand, feels resistance from a block, pushes harder, and then perceives the movement of the block. “Cause” is then mapped to such an experience later when the child maps the term to physical interactions s/he has experienced and/or observed. And with this comes a chronological direction of force (or causality) in the conceptualization. Causation has a direction indicating the source of the supposed cause toward the effect. Hand away from the body “pushing” the block; hand and block “pulling” toward the body, etc. Pushing and pulling then become fundamental aspects in the conceptualization because they are born out of one’s earliest activity in the world. Again, this is a leap, but they believe it explains how one can arrive at a concept of causation empirically, and the continual use of push/pull relationships in conceptualizations of causal relationships supports the notion that it’s grounded in embodied activity.

Finally, Glenberg and Kaschak also claim there is enough evidence to believe language about abstract phenomena is grounded in bodily activity. To explain why, they consider a study conducted by Ochs, Gonzales, and Jacoby (1996) regarding the interactions of scientists discussing high-energy physics. The scientists were discussing changes in the temperature of a substance and the resulting domain state, depicted on an x/y graph. When describing changes in temperature, the scientists used arm gestures to simulate the changes in temperature. When discussing a rise in temperature, the speaker would raise an arm; when discussing a drop, the speaker would lower an arm. However, the arm movements were not ubiquitous—they only seemed to accompany speech when the scientists were having trouble communicating and understanding a new hypothesis. The researchers observing this believed the bodily movements were allowing the speaker(s) to make the abstract hypotheses concrete via bodily movements and grounding them in sensorimotor states. Such behavior is acceptable and even necessary with the indexical hypothesis because it allows both speakers and observers to gain understanding of abstract concepts being presented, via meshing. In the scenario of the scientists, the goal of speakers and listeners was to plot the substance on an x/y graph. Thus, the up or down movement of the arms of the speaker allowed all involved to grasp how the resulting concept applied to the rise or fall of the substance's temperature—since their aim was to depict the rise and fall by changes in vertical positioning on the graph.

Similar studies regarding gesture have observed similar behaviors. Barbara Tversky (2009) describes an experiment in which subjects were divided into two groups: one with pencil and paper, one without. The groups were then asked to solve a variety of abstract problems. The interesting overlap was that gestures in the group without pencil and paper became common for the same problems in which the other group relied on pencil and paper. Once again, it appeared that processing certain abstract notions is more effective when one can incorporate bodily movement or rely on other observable explanations (such as drawing or writing it out).[23] In another experiment conducted by Francis Rauscher, Robert Krauss, and Yihsiu Chen (1996), it was observed that when subjects' gestures were restricted, their speech was slower and filled with more pauses in unnecessary places than normal (as opposed to at the end of a sentence or some other grammatical clause). The researchers believe the delays became more prevalent as gestures became limited because gesturing helps speakers find the words to communicate more effectively. The bodily movement allows one to simulate the concept in order to recall it quickly. This particular study was tied less to the essential nature of bodily movements in language processing—as it involved only spatial concepts, and the speed of recall rather than the introduction or utilization of novel concepts—yet it, along with several other studies (Ehrlich,

[23] Tversky also uses this experiment to support the notion of external cognition, saying the findings "suggest that gestures were used to off-load and organize spatial working memory when internal memory was taxed." (Tversky, 2009: 210)

Levine, and Goldin-Meadow, 2006; Goldin-Meadow, 2003; and McNeill, 2005) indicate that bodily movements, at minimum, boost the cognitive efficiency of speakers and, at maximum, play an essential role in the construction, storage, and retrieval of linguistic concepts.

This last claim, however, is far too optimistic at this point, and not simply because it is a difficult claim to examine via empirical research. The philosophical arguments that have provided the foundation for such research have not yet been discussed. After all, it's not as if these experiments were conducted in a hypothetical vacuum wherein the goal was pure observation. Each of the experiments described above were tailor-made to test very specific notions of language and its relationship to the body or modal processing. The experiments conducted and conclusions reached by Glenberg and Kaschak, Tversky, and others has been influenced greatly by the work of Lakoff and Johnson on the fundamental role of body-based metaphors in human language. Therefore, in order to consider the validity of these physically grounded explanations of abstract language, and appreciate the impetus for their existence, it is crucial to look at the theory advocated by Lakoff and Johnson.

The Theory of Metaphor

To put it concisely, Lakoff and Johnson's theory is vital to the story of embodiment because rather than considering how one thinks in isolation from experience, their account ties *how* one thinks to *what one is able* to think and think about. Their primary claims are: (1) Metaphors provide the boundaries—expansive and restrictive—for one's cognitive capabilities and, (2) the metaphors we live by are based in large part on the particular types of bodily experience of which human beings are capable.

Metaphor, though a tool in language, is not really about language; it's about grounding concepts in the manner described earlier—providing a foundation from and through which concepts become meaningful. As Lakoff says, "the locus of metaphor is not in language at all, but in the way we conceptualize one mental domain in terms of another." (Lakoff, 1993: 203) An effective metaphor will allow one to grasp a given concept by connecting it to concepts already understood in another domain. The previous sentence is an example. One cannot literally 'grasp' a concept. Literal grasping involves the physical act of wrapping one's fingers and palm around an object. But the notion of holding on to something tightly to the point that it is under one's control (such as grasping a baseball) or at least near enough to be used in some way (such as grasping a railing) can be applied to concepts. In this way, the metaphor of grasping allows one to understand 'understanding' better—understanding a concept is like grasping an object. According to the theory of metaphor such cross-domain mapping occurs across concepts used in everyday language such as time, states, change, causation, purpose, and more.

In *Metaphors We Live By* (1980), Lakoff and Johnson lay out a variety of such metaphors to prove the ubiquity of this occurrence in abstract concepts and explain how this allows one to understand and experience any given thing in terms of another. Some examples include:

- Theories are buildings—they need support, they have foundations, they need to be constructed, and they can collapse.
- Ideas are food—they can be half-baked; they can leave a bad-taste in one's mouth, they might ferment for a while; and they are sometimes food for thought.[24]
- Love is a physical force—it generates electricity; there are sparks between lovers; people gravitate toward one another.

---

[24] Ideas can also be people (given birth to; spawned from another idea; etc.), plants (come to fruition; have branches; etc.), products (generated; smoothed out; etc.), commodities (they're valuable and can be bought or sold), resources (they can run out; be wasted; used up; etc.), cutting instruments (they're sharp and incisive), and fashions (they go out of style; they're outdated).

- Wealth is a hidden object—one seeks a fortune; a person can have newfound wealth; a person can be a gold digger.

- Eyes are containers for emotions—one can see the fear in another's eyes; eyes can be filled with anger; love shows in one's eyes; eyes become welled with emotion.

- Emotional effect is physical contact—a death in the family can hit someone hard; someone can be a knockout; one can be struck by another's sincerity, or touched by another's remark.

One could look at these examples as interesting, but not particularly valuable in terms of establishing the means and method of cognition. A skeptic could say, "Sure, metaphor expands one's understanding in that it allows one to utilize previously understood concepts in new ways—like assembling Lego blocks in various ways—but that doesn't tell us how the blocks come to be in the first place." As stated before, Lakoff and Johnson believe the metaphors are responsible for grounding abstract concepts because the metaphors bind meaning to physical concepts and possibilities. The abstract concepts are grounded by metaphors because the metaphors are derived from direct physical experience. One physically grasps an object with a hand and, due to a process akin to the meshing proposed by Glenberg, forms the linguistic concept of GRASP, which is then tied to the bodily experience. Thus, when one constructs[25] the abstract notion of understanding with the property (and concept) of grasping, the concept UNDERSTANDING is not based on relationships between symbols alone. Instead, its meaning can be reduced to physical experiences.

The dependence on metaphors and the metaphors' reducibility to physical experiences then provides the boundaries for language and conceptualization of any given concept. As the metaphors are based on actual or potential bodily experiences, concepts are bounded by the actual or potential experiences our particular bodies can have. Just as Varela, Thompson, and

[25] Another metaphor!

Rosch describe color experience as both generated and limited by the particular visual system of human beings, experiences of every kind will be conceptualized in terms specific to the sensorimotor parameters of our species. This, for example, is why one can have concepts like LEFT and RIGHT. If human beings had several more appendages protruding from the center section of our front and back, such concepts could not exist as they do now. FRONT and BACK are also paradigm examples. Since humans' visual perception is situated on one side of the body, we are capable of categorizing other objects and bodies in a similar fashion and then applying this physical knowledge to generate abstract terms such as "Frontrunner" or "Head of the class" where all others are implied to be "in back" or "behind."

This can also be observed in terms and concepts that have evolved as humans have gained greater knowledge of human physiology. A person can be described as the "backbone" of an organization, implying their importance for the organization to function properly. Someone else might also be the "brains" of the organization, implying importance for ingenuity, planning, and direction. A disruptive person or persons may be called a "cancer," upsetting the operation as a whole and possibly negatively affecting others with whom they come in contact. Likewise, someone can "remedy" a bad situation.[26]

And, aside from the ways in which metaphors are utilized for concepts used within language and cognition, they are essential to understanding "mind" at all. In regards to traditional cognitive science, the notion of a formal language of Mentalese allows one to comprehend cognition as if it were similar a natural language. But, as Lakoff and Johnson point out, natural languages have phonetics, phonology, and morphology, whereas formal languages do not.

---

[26] Perhaps one of the greatest, extended examples of this is Thomas Hobbes' depiction of the commonwealth in *Leviathan*. Beginning with the illustration on the cover and throughout the text he continually describes his ideal commonwealth as a body, complete with bodily metaphors about properly functioning organs, possible sicknesses, and the keys to healthfulness.

"Intonation in natural language is on a separate plane from phonetic segments; that is, an intonation contour typically extends over many segments and can have meaning separate from the meaning of the segments." (Lakoff and Johnson, 1999: 259) A subtle change in pitch, a slant of up-speak at the end of sentence, a flat, droll delivery of a spoken word—these intonations can imply sarcasm, skepticism, or doubt regardless of the syntax or semantics of the language. Similarly, the expressions that can be affected by intonations in natural language can have multiple meanings. "Yard" can be an outdoor space around a structure or a unit of measurement; "fall" can be a season or a verb; etc. The reason for pointing this out is to help reveal the value and flaws associated with metaphor. A formal language of thought is described as a language, relying on one's understanding of natural language; and natural language is described as if it relies on a more fundamental formal language. This can lead persons to conceptualize both formal and natural languages as if they are related. However, "natural languages are not subcases of formal 'languages,'" for Lakoff and Johnson, "a formal 'language' is therefore only metaphorically a language." (Lakoff and Johnson, 1999: 260) So, the metaphor that allows one to understand the concept of a formal language leads to a false sense of understanding about the nature of formal languages; the two are not as similar or interrelated as supposed.

Just as the formal "language" of the mind may be misconstrued as a language due to metaphor, the mind itself can fall prey to the inadequacies of metaphor. Notions such as "the mind is a computer," *inner* thoughts, or mental *space*, portray the mind as if it is a storehouse which clearly delineates the external world from the supposedly internal mind.

> We conceptualize the mind metaphorically in terms of a container image schema defining a space that is inside the body and separate from it. Via metaphor, the mind is given an inside and an outside. Ideas and concepts are internal, existing somewhere in the inner space of our minds, while what they refer to are things in the external, physical world. This metaphor is so deeply ingrained that it is hard to think about mind in any other way. (Lakoff and Johnson, 1999: 266)

This, they believe, assists the traditional model and hampers theories of embodied cognition. It is much easier to rely on container-like metaphors because experience of containers is available. But if the mind is embodied it is not and cannot be experienced in the same removed, third personal manner; it is inseparable from interactions with the world and "is part of the very structure and fabric" of these interactions. If correct, then there is no legitimate metaphor that corresponds to what a mind is like.[27] Thus, one relies on physical object metaphors and begins to develop a concept of the mind that possesses object-like properties. But this does not get one any closer to an understanding of the actual nature of the mind.

This is not to disparage all uses of metaphor when discussing the mind; like it or not, metaphors are unavoidable when arguing about the mind. Coupled, sandwiched, language-like, computer-like, or whatever, in order to discuss the mind it appears that it must be presented in terms which are applicable to other objects or experiences. Lakoff and Johnson are basically demanding that, at minimum, one recognize how metaphors are shaping notions about the mind and negatively affecting one's ability to discuss (and accept) the theory of embodied cognition. But this minimal claim is not their goal. They support a fully embodied conception of mind that entails an inseparable relation between language, modal perceptual systems, and sensorimotor affordances.

The role of and reliance upon language in an embodied account, however, extends well beyond metaphor and a few experiments performed on human beings. Assuming Lakoff and

[27] This is also a problem for the notion of "coupling." As mentioned in the section on coupling, this term is used to connote a sort of interconnectedness between mind, body, and world, but even this metaphor implies separable properties between the three that may not be present in an embodied cognitive system.

Johnson are correct in their claim that concepts of the human mind are impeded by metaphor since it is part of the very fabric of the discussion, then our discussions—whether based on empirical studies of humans or phenomenological analyses of language—may always come up short in terms of definitive conclusions. Yet, if we could achieve a more third-personal perspective then we would at least have a chance at surmising a more objective account of cognition. Clearly, this can't happen in regards to oneself or (for the foreseeable future) in another human being. Our minds and others' minds are already functioning by the time any analysis or experimentation occurs. But the realm of artificial intelligence offers a much different opportunity.

With AI, we are the architects, engineers, programmers, and caretakers of (potentially) cognitive systems. And though bound by our language in terms of assembly instructions and descriptions, they can operate on input and processing languages radically different from our own. Therefore, in the next section I will look at embodiment in artificial intelligence and how some researchers and developers believe a coupled cognitive architecture is essential to realizing a successful artificial intelligence.

3.5 Efforts in artificial intelligence support embodiment

The challenge of attempting to develop a comprehensive theory of one's own cognition while participating in and being limited by one's own cognitive capabilities is lost on no one in the field of cognitive science. Whether via a phenomenological analysis of one's personal experience or empirical studies of other conscious subjects, there is no denying that observations and theories of cognition by what's already there—in terms of one's own cognitive skills and the organic makeup of every other cognitive being—make it extremely difficult to provide a non-circular account of cognition. Awareness of this challenge has inspired many to seek insight into and justification for certain theories outside of the "natural" or organic realm and into the arena of artificial intelligence (AI). The appeal of AI is also its greatest difficulty: creating pseudo-cognitive systems from the ground up. The potential upside is that it offers researchers, programmers, engineers, philosophers and others the opportunity to *construct* a cognitive system—not simply detect it, observe it, study it, and theorize about it, but make it. This means building the perceptual systems, determining the path and categorization of input, inventing a novel language of thought (if necessary), as well as writing the rules and/or algorithms that govern courses of action. As the last two items in that list reveal, many of the properties of cognition assumed under certain theories can be included or excluded in the formation of the software and hardware in order to see what works best, and which, if any, are necessary or sufficient for successful cognition. Over the last several decades some efforts have been based on more traditional computational theories, while others have incorporated a more embodied approach. Though no one has yet to create an AI capable of fully simulating human cognition—and this is the downside—both traditional and embodied approaches have made significant strides. However, embodied approaches seem to be yielding AIs much more adept at dealing with the continually new and changing situations humans face every day relating to our embedded, embodied interactions with the world.

Traditional Cognitive Science Approaches

During the 1950s and 1960s, Herbert Simon and Allen Newell gained a good deal of attention for creating the General Problem Solver (GPS). The GPS—a novel computer program intended to solve logic problems in a manner similar to humans—was the culmination of a rigorous line of research aimed at uncovering "how particular human behaviors come about…[and] the mechanisms that enable them." (Simon and Newell, 197: 146) Thus, they wanted to go beyond mere calculation and create a program that could traverse a rational landscape in order to determine the best course of action to solve a rational puzzle. This was very different than calculations wherein there is usually a single equation to be followed. The logic puzzles, though they had a single correct answer, could be solved several different ways. And when I say the GPS was the result of rigorous research it is because Simon and Newell followed a comprehensive multi-step strategy in pursuit of this lofty goal. It is worthwhile to look at this strategy briefly, as it provides a thorough explanation of the steps and potential benefits of examining cognition through AI. The strategy goes as follows:

1. Discover and define a set of processes that would enable a system capable of storing and manipulating patterns to perform complex nonnumerical tasks, like those a human performs when s/he is thinking.

2. Construct an information-processing language, and system for interpreting that language in terms of elementary operations, that will enable programs to be written in terms of the information processes that have been defined, and will permit those programs to be run on a computer.

3. Discover and define a program, written in the language of information processes, that is capable of solving some class of problems that humans find difficult. Use whatever evidence is available to incorporate in the program processes that resemble those used by humans. (Do not admit processes, like very rapid arithmetic, that humans are known to be incapable of.)

4. If the first three steps are successful, obtain data, as detailed as possible, on human behavior in solving the same problems as those tackled by the program. Search for the similarities and differences between the behavior of program and human subject. Modify the program to achieve a better approximation to the human behavior.

5. Investigate a continually broadening range of human problem-solving and thinking tasks, repeating the first four steps for each of them. Use the same set of elementary information processes in all of the simulation programs, and try to borrow from the subroutines and program organization of previous programs in designing each new one.

6. After human behavior in several tasks has been approximated to a reasonable degree, construct more general simulation programs that can attack a whole range of tasks—winnow out the "general intelligence" components of the performances, and use them to build this more general program.

7. Examine the components of the simulation programs for their relation to the more elementary human performances that are commonly studied in the psychological laboratory: rote learning, elementary concept attainment, immediate recall, and so on. Draw inferences from simulations to elementary performances, and vice versa, so as to use standard experimental data to test and improve the problem-solving theories.

8. Search for new tasks (e.g., perceptual and language tasks) that might provide additional arenas for testing the theories and drawing out their implications.

9. Begin to search for the neurophysiological counterparts of the elementary information processes that are postulated in the theories. Use neurophysiological evidence to improve the problem-solving theories, and information from the problem-solving theories as clues for the neurophysiological investigations.

10. Draw implications from the theories for the improvement of human performance—for example, the improvement of learning and decision making. Develop and test programs of application.

11. Review progress to date, and lay out a strategy for the next period ahead. (Simon and Newell, 1972)

As should be clear from the first few strategies, Simon and Newell were basing their approach on a thoroughly traditional method, presuming pattern manipulation and an information-processing language in order to achieve any sort of problem-solving ability or behavioral output. And though they admit pursuing these strategies only up to step #8 (until to 1972), steps #9–11 have been pursued by more current efforts, and the overall process remains applicable in utilizing AI to research cognitive theories.

The GPS they created by following this strategic path provided important, yet limited, insight into cognition. Perhaps most importantly was the “general” part of the General Problem Solver. The GPS was not the first computer program designed to solve problems, but it was the first designed to have a general reasoning capability which could be applied to novel situations rather than simply being able to perform tasks with specific goals in mind. Two notions were essential to the endeavor: (1) cognition is primarily a process of symbol manipulation, and (2) humans often engage in means-end analysis when solving problems. The first provides justification for pursuing programmable AI at all, since the methods at the time involved defining rules for input/output processing via symbolic programming code. The rules created by Simon and Newell, however, were a great step forward because they were not simply equations the machine would use to process input. Instead, they were heuristics that would allow the GPS to solve problems by navigating a goal-driven “problem space.” According to Simon and Newell, means-end analysis (human or AI) operates in a cognitive problem space in which there is the current (or initial) state one is in, the goal state in which one aims to be, and all the possible states that exist between the two. Hence, part of their challenge was to create this problem space for the GPS and then the rules that would allow it to go from the current state, through various possible states, and finally to the goal state.

Initially, this is just an abstract consideration because the actual problem space of any given problem-solving system depends on the physical components of the cognitive system and the rules governing its cognitive processes. For example, an artificial system tasked with providing data on air temperature will go through very different problem spaces depending on if it has a thermometer attached to it and, if so, which type of thermometer, the materials that make up the thermometer, how the thermometer's readings are symbolically encoded, if the system is expected to provide information in Fahrenheit or Celsius, etc. However, at least in regards to human cognition, Simon and Newell believe the problem space is represented and determined by "a few, and only a few, gross characteristics of the human information-processing system [which] are invariant over task and problem solver." (Simon and Newell, 1972: 148) This could be something such as the rules and function of the language of thought. The assumption is that the underlying language of thought remains the same, regardless of the external environment or changes in the non-neural physiology of the person. Thus, rather than focusing on the entire human information-processing system—much of which may not be replicable in AI, such as neural structures and electrochemical processes—they aimed to observe and uncover fundamental aspects of rational thought that remain constant regardless of the task or situation of the person, and then replicate these in the GPS. The hope was to that human-like cognition could be achieved in an artificial cognitive system that could simulate and function within a rational, means-end problem space.

As far as Simon and Newell's assumptions and goals, the GPS was fairly successful. To understand how it could be judged as successful, it's important to consider their method. Since the goal was to *simulate* human cognition, they first needed to establish a control example of cognition. To achieve this they studied the paths human subjects took in order to complete certain cognitive tasks. They asked human subjects to solve various logic problems—such as transforming R & (~P ⊃ Q) into (Q v P) & R—while talking through his/her reasoning in order to get a consistent and clear picture of the human problem space. Requiring human subjects to first talk through the problems allowed the researchers to ascertain the rules that would be applied by the GPS (steps #4-6 from the strategy mentioned earlier). Though different human subjects went about solving the problems in slightly different ways, Simon and Newell were able to identify a general path and translate it into the GPS's cognitive language. And it worked well. Logic problems, such as the one above, were given to the GPS by providing the initial state—R & (~P ⊃ Q)—and the end or goal state—(Q v P) & R—and then allowing the program to traverse the problem space between the two in the manner it deemed best. Based on the programming (and without going into painstaking detail as to how the GPS operated) the GPS followed a similar protocol as human subjects and solved the problems accurately.

The fact that the GPS was able to perform along lines similar to human reasoning lead Simon and Newell to suggest "GPS provides a rather good approximation to an information-processing theory of certain kinds of thinking and problem-solving behavior. The processes of thinking can no longer be regarded as completely mysterious." (Newell and Simon, 1961: 2016). With the concept of a computational means-end problem space created and put to the test, Simon and Newell believed they had found strong empirical evidence that such processes play a fundamental role in human cognition. And though a great deal of human cognition outstrips the rules and symbols required to solve formal logic problems, this at least indicated that some (supposedly) uniquely human cognitive capabilities could be replicated by applying the right symbolically-encoded rules to a well-defined set of possibilities and goals in an artificial agent.

Following the GPS, the remaining challenge was to see just how far this model could go in simulating cognition and (as stated in steps #6-9) to find neurophysiological counterparts to even more complex problem-solving programs designed along this theory. However, in the decades since, even with more complex problem-solving programs developed along similar lines, conclusive evidence for ubiquitous computational means-end problem spaces in human cognition has yet to be found. This is not to say there isn't evidence that human beings solve problems in such computational ways. There are instances where humans consider various rules and then select and apply them in order to achieve a desired goal. However, even when doing so, embodiment theorists say one usually does so in less computationally abstract ways because our cognitive "programming" is inherently embodied and embedded.

The GPS exhibited the rational capability to solve an abstract logic problem. But the rational challenges facing human beings are rarely, if ever, so free from the interactions and entanglements with the world. At a minimum level, one would receive the input or challenge of the logic problem via some sensory perception. One might hear the problem, see it written, feel it in braille, etc. So, from the outset the abstract consideration is intertwined with sensory data. Additionally, a human subject is likely to ask "Why?" Why solve the problem at all? What goal will be pursued or furthered? Is it for a grade that will help or hurt my degree? Will it impress my boss? How will I feel if I solve it or fail to? And so on. The point being that our means-end reasoning rarely places the end in a completely abstract realm. The goals generally have tangible consequences, even if the tangibility is a sense of accomplishment or the esteem of another.[28] Many embodied theorists believe this to be the result of our inherently embodied and embedded cognitive faculties. Together, "embodied" and "embedded" imply that one cannot explore and express one's cognitive abilities removed from one's body or the world in which one finds

[28] I keep using vague terms about embedded and embodied, such as how we often or generally have non-abstract goals, because I am aware this is a highly debatable claim. However, like many issues, it is not within the scope or purpose of this discussion to uncover the tangible or abstract nature of human reasoning. Thus, I am phrasing this so as to allow for the caveat that there are arguments in favor of less tangible reasoning without detailing the pros and cons of such arguments.

oneself. Additionally, it may be impossible to partition off the cognitive programming from the body and/or world. This may sound like a simple restatement of the idea of coupling already discussed, and in a way it is. But the issue now is how this affects successful AI. According to some researchers and theorists, where Simon and Newell's model finds its limits is where embodied and embedded approaches excel.

Embodied and Embedded Programming

One criticism of the GPS and any conclusions of cognition drawn from it, or a similarly constructed AI, is that it is disembodied in very significant ways. A program capable of solving logic problems could be considered minimally embodied, in that it must have an input system where it can "perceive" or receive the problem and an output system where it can communicate the solution—which could be achieved with just a screen and keyboard setup.[29] But, based on this programming alone, critics will say the GPS could not traverse the world, could not use limbs to generate locomotion, could not process novel input, could not see novel environments or objects and determine if and how to interact with them. Or at least could not accomplish these tasks in a manner that bespeaks of human-like cognitive abilities. Of course "could" or "could not" can be viewed as weak and ambiguous statements. Perhaps an AI developed along Simon and Newell's strategy, exemplifying the traditional model of cognition, isn't able to do these tasks because computing hardware and software have yet to reach the complexity and speed of human cognition. But, once hardware and software are able to achieve a similar level of computation, the AI would be able to perform much more elaborate rational tasks that could

[29] This may not even be the most minimal embodiment. If the problem(s) can be solved by mapping the right answers to the questions, then the form of embodiment (whatever it might be) only needs to allow the system to solve the problem and confirm or know that it has done so.

include logic puzzles as well as problems related to interpreting and traversing an ever-changing external environment—problems such as climbing stairs in an environment of shifting light sources, identifying and avoiding and/or catching falling objects or varying shapes, traversing unstable surfaces, staying balanced on slippery surfaces, etc. Though these examples don't obviously require the rationality that the GPS exhibited in solving logic problems (and could therefore be seen as different concerns in problem solving and AI), the critique is that a truly *general* problem solver must be able to display the traits of the GPS in traversing the logical problem space, as well as the ability to solve problems faced in the physical world by all embodied agents.

Along this line of criticism, it's correct to say the GPS and AIs designed on similar principles can't currently do many cognitive tasks. But this doesn't entail that they couldn't with greater strides in computer programming and design. Maybe with faster processors and better sensors and motors, a machine designed on traditional principles of cognition could successfully complete such tasks. For this reason, and why AI is worth discussing in this context at all, several embodied theorists say that AIs based on traditional computational models will never be able to achieve human-like interpretations of and interactions with the world. Even with faster processors, motors, sensors, and the like, AIs designed along those lines will always fail. The reason for this perpetual failure, they say, is due to the assumptions underlying traditional—or classical—AI. Instead of relying on the classical approach, the system or agent must be designed based on embodied and embedded approaches that couple the internal rules and processes with sensory systems and the external environment in ways that are more dynamical than computational.

As mentioned in the earlier discussion of coupling, dynamical systems are simply systems whose behaviors evolve or change over time. More explanation will follow, but for now contrast this with the GPS. With the GPS, the problem space and the rules governing it had to be clearly defined and programmed in advance. And they did not change over time. Though the logic problems posed to the GPS varied, and possibly even how it chose to solve each problem, the rules and the possibilities open to it did not. It engaged with the input, but statically rather than dynamically. The internal system was not coupled with anything outside of it, so a change in input would not result in a change to the organization or fundamental operation of the GPS. Literally and philosophically, the GPS was disembodied and only needed its internal processes and capabilities to complete its cognitive task(s). To understand why some researchers view this as an unacceptable way to achieve human-like cognition, consider Rodney Brooks' attempt to apply a similar programming model to a mobile robot with a greater range of perceptual and motor abilities.

Not long after Simon and Newell created the GPS, Brooks constructed a robot, Shakey, that was the first designed to visually interpret its environment. It did so by demonstrating a traditional cognitive science approach to perception in artificial intelligence. While successful in many ways, Brooks would later criticize Shakey as a failure.[30] Standing about six feet tall and outfitted with a TV camera, a wireless video system, a triangulating range finder, and bumbers, Shakey looked like a supped-up box on wheels. Similar to the GPS, once on, Shakey was designed and programmed to reason about its own actions by determining how to navigate its surroundings via means-end reasoning that traversed a problem space. In many ways, Shakey

[30] Some of the criticism is in the name—"Shakey" refers to how the robot moved when attempting to complete the tasks given to it. It never achieved the smooth motion of a being with proprioception. Instead, it jerked around somewhat clumsily. But, then again, it was just the first in a long line of efforts to create robots that could move about fluidly.

was groundbreaking. The robot could navigate its environment to complete a given task and determine the best manner for doing so. In one task, an operator would type the command to push a block off a platform. After receiving the input, Shakey would look around the room, identify the platform with the block on it, and locate a ramp that would allow the robot to reach the platform. Then, the robot would push the ramp into position near the platform, roll up the ramp, and push the block off the platform. Completing the tasks might take hours or even days, but the gigantic accomplishment was the fact that Shakey could do these things somewhat autonomously.

In terms of cognition, Shakey was designed to replicate symbol manipulation in perception and action. The command from the operator played the role of symbolically encoded goals of the robot, which would then prompt focused, selective processing of visual input that would then be encoded by Shakey's internal processing unit, checked against concepts already encoded within the robot's "mind" (such as BLOCK and PLATFORM), which would then prompt further physical actions that would provide more input, and so on. This was a sense→model→plan→act path of cognition, according to Brooks. The physical form of Shakey was not considered essential to the robot's cognitive abilities. So long as the CPU was receiving input and able to send signals that could produce action, the artificial intelligence was functioning.

Similar to a brain-in-a-vat, one could imagine Shakey's programming working effectively in a fully digital environment wherein the blocks, platforms, and walls, as well as the shape and size of Shakey, were nothing more than computer-generated shapes in a computer-generated environment. Shakey could still "see" and "move" even though the visual perception and movement would really be nothing more than the translations of digital symbols. Though this task, and the successful completion of it, is far from illustrating the myriad mental states and propositions of human cognition, Shakey at least appeared to represent a working example of how a visual perception and action could operate according to the tenants of traditional cognitive science. The symbol manipulation successfully resulted in physical interaction with the world, even though the symbols were grounded in abstract programming rather than physical reality. So why did Brooks eventually believe Shakey to be a failure of artificial intelligence?

Part of the problem with Shakey was in the experiment itself. Shakey was designed to handle a very limited set of circumstances in a very unrealistic environment. Unlike the world most humans traverse, the robot's environment was always static, well lit, and objects were clearly delineated from other objects—nothing was going to move without Shakey moving it, and nothing was going to change without Shakey changing it. In contrast, the world we normally find ourselves in is much more difficult to decipher and in a near constant state of flux (sometimes regardless of one's actions and sometimes due to one's actions). Additionally, the manner in which we categorize perceivable objects is in flux due to differing goals when interacting with the world. This can be observed in how a stool can be categorized as an object to stand on when one is trying to reach a higher height, or as something to sit on when one is tired, or as a place to set other objects upon when rearranging a room, etc. It can also occur rapidly and unexpectedly if, say, a person stumbles while walking by a chair. The chair may be perceived and categorized initially as an obstacle to walk around, but when stumbling it may suddenly be utilized as an object with which to regain one's balance. Nothing in the chair has changed, but one's perception and/or categorization of it has changed due to embodied action.

Connected with this example of stumbling is also the issue of time. As just described, when stumbling, an object is quickly re-categorized or utilized in a novel way. However, such an event may occur so quickly that categorization itself does not occur. That is, there may not be enough time to perceive, conceptualize, and categorize before acting; not enough time for the sense→model→plan→act path of cognition. Instead, a more situated model of computation and cognition may be taking place. “Situated” in the sense that the model must tie together the cognitive processes with the form of the agent (embodied) as well as the environment with which the agent’s body is interacting (embedded). For Brooks, this all provides reasons to question the success of Shakey. Shakey represented a huge step forward in robotic programming, yet the very limited aspects of its success (means-end tasks in a limited, static environment) revealed how much more robust an artificial intelligence must be in order to replicate human-like behavior in an ever-changing world.[31]

Though Brooks and others with similar views believe there is value in the traditional model in terms of cognition and AI (or “physical symbol-system hypothesis,” as he calls it), one of the main problems hindering AIs like Shakey or the GPS is that “the physical symbol-system hypothesis cannot constitute complete explanations of intelligence, precisely because they abstract away the details of symbols’ implementation.” (Brooks & Stein, 1994: 10). Designing a robotic AI along the traditional path is like developing a brain-in-a-vat first, complete with its own internal language of thought—with rules governing processes based on the characteristics of the physical components that make up the system as well as the encoded data it can hold and communicate with—and then engineering the best way to connect these systems to sensors, pistons, limbs, etc., that will form the body. But, as has been mentioned several times, in this

---

[31] This wasn’t necessarily due to the researchers inabilities to foresee the limitations of Shakey. Brooks has said the goals and efforts of artificial intelligence design at the time were determined more by the technological resources available to researchers rather than the researchers themselves. (Brooks & Stein, 1994) With the advent of much faster, more sophisticated, dynamic processors, Brooks and others began to see the possibilities of programming that did not rely on a single, central representation of the world.

approach the symbols are grounded in the programming. It is already determined how light with wavelengths that correspond to the color red will be encoded, for example, as well as sounds that correspond with dogs barking. Those physical events still have to be detected by the robotic AI in order to be perceived and encoded, but the symbols that will come to represent them do not depend on them to exist. To have the meaning (or play the role) of redness or of the sound of a dog bark the symbols depend on the external event for certain semantic information, but the underlying potential for what role the symbols can play is inherent to the symbols themselves, based on the physical characteristics of the material and hardware, and the rules of the programming. In this way, as the continuing refrain from the traditional theorist says, the semantic *value* of the symbols may be constituted by the external stimuli, but the ability to have functioning symbols is not.

For Brooks, this is not the best way to approach cognition, artificial or organic. As he says,

> In order for a brain-in-a-box to connect to a body, all symbols must be derivable from sensory stimuli; but in addition, there are portions of the system…that cannot be seen from the symbolic side of abstraction. Thus, while symbolic approaches to cognition may provide us with tremendous insight as to how intelligence might work once we have symbols, it can neither tell us how to construct those symbols nor assist us in the identification and manipulation of the non-symbolic portion of our system. (Brooks & Stein, 1994: 11)

The assumption here being that in order to create an AI capable of exhibiting more humanlike behaviors and cognitive abilities that can deal with novel situations—rather than a robot with the processing capabilities to recognize a very controlled, finite set of patterns in the world and correlate those with an equally well-dictated set of responses—requires that the robot's abstract symbols and semantics be grounded in its physical interaction with the world rather than its internal language and programming alone.

To remedy the problems of the traditional approach and apply these notions of embodiment and situatedness, several designers and engineers (including Brooks) have applied dynamical systems models to AI. One essential aspect to such a model is including the agent's "situatedness" when creating a model of cognition. Being situated entails an agent having a body that can be placed in an environment and that this body can interact with the environment in which it is placed. Thus, in terms of programming, situatedness demands more factors than the disembodied programming and language of thought.

This is where the *dynamic* nature of the model comes in. Dynamical systems change over time. Thus, if the mind-body-world model is to be dynamic, some aspect of it must evolve, vary, or alter as it is functioning. In short, there must be activity wherein the agent is interacting with the environment and this must feedback into, and affect, the rules and goals defining the programming. The program and processes of the AI become situated in that they cannot be disconnected from the sensory mechanisms that make up the physical form of the AI, nor from the information received through those mechanisms. This is one way dynamical models differ significantly from classical AI and traditional cognitive science: situated activity is an essential component. But this needs to be defined more clearly.

According to Randall Beer, situated activity in dynamical systems has three primary components (Beer, 2014):

1. Concrete action—actually taking action in the world is more fundamental than the abstract descriptions that we sometimes make of it. While conscious deliberation clearly has its role, the ultimate job of an intelligent agent is to *do* something, to take some concrete action with consequences beyond its own skull.

2. Situatedness—an agent's immediate environment plays a central role in its behavior. This environment is not only a rich source of constraints and opportunities for the agent, but also a context that gives meaning to the agent's actions.

3. Interactionism—an agent's relationship with its environment is one of ongoing interaction. The environment does not serve merely as a source of isolated problems for the agent to solve, but rather a partner with which the agent is fully engaged in moment-to-moment improvisation.[32]

To understand what this means in terms of AI design, consider Beer's method of addressing categorical perception.

Categorical perception refers to the phenomenon whereby continuous changes in a distal stimulus produce discontinuous changes in perceptual states.[33] In a very broad sense, this is why one can recognize green from black, the street from the trees, and leaves from squirrels. In a more narrow sense, this is how one is able to perceive light of different (though very similar) wavelengths as being the same shade of a color, while light of more dissimilar wavelengths are perceived as different shades of the same color (or different colors altogether). For example, when looking out my window I am able to see multiple shades of green grass. Some lighter, some darker, yet all categorized under GREEN. My desk and laptop touchpad are both smooth, but the desk is slightly less smooth than the touchpad. A strawberry is sweet, but cotton candy is sweeter. This ability to perceive categorically, both broadly and narrowly, is fundamental in cognition as it acts as the groundwork for developing the myriad concepts one can possess. Without categorization and subcategorization, perception would just be a buzzing blur of confusion.

The importance of categorical perception has been recognized for centuries, but attempting to create this ability in an artificial perceptual system is relatively new. In the case of

---

[32] Beer credits Heidegger as providing the philosophical roots to these claims. In particular, Heidegger's famous example of the hammer: one encounters the hammer as a hammer to achieve a certain end (to drive a nail), and when in the act of hammering the hammer ceases to exist on its own, in a sense, and becomes an extension of one's arm, and only reemerges as a separate entity when it suddenly breaks or one takes an intellectual attitude toward it.

[33] My thanks to Jack Lyons for providing this nice, concise definition of categorical perception.

Shakey, this was done in a way that did not require much overt activity other than the activity of visual perception and having an already existing internal framework that could distinguish a limited set of possible percepts in one way or another. The vagueness of this last statement is not intended to downplay such ability.[34] The "one way or another" is a terrifically difficult issue to address in designing and programming the AI. Shakey's camera provided visual input of the objects surrounding the robot and the internal processes distinguished the objects, categorized them, and then used the same visual system to determine the best course to take in order to complete a given task. However, as mentioned before, the objects did not move unless Shakey moved them, and Shakey was only programmed to recognize the objects necessary to complete the task. This is not indicative of many of the perceptions one encounters in an ever-changing, uncertain world. For example, Shakey could have been asked to push a block off a platform that was moving, or would be moved in a matter of seconds. Or there could have been other robots attempting to knock it off first, or some of the ramps could have been broken, or changing lights could have made it difficult for Shakey to distinguish ramps from other objects, or the ramps could have been on casters and blown around the room by strong fans. Shakey—or an AI with similar programming—could still have successfully pushed the block off in these more dynamic scenarios, but to do so would have required much faster, more nimble, more advanced categorization abilities. Faster because the environment would not sit still for hours or days waiting for Shakey to go from point A to point B and complete task C; and more nimble and advanced because it might require categorizations broader and less definite. If some ramps had unseen casters, then they could be categorized as ramps initially (with the property of being

[34] Kant famously proposed an a priori categorization schema for perception and cognition, and debate continues on the nature and origin of this cognitive power. Whether such categories are ontological, realist, or descriptivist is not necessary for this discussion, so I will assume categories exist in cognition without debating the metaphysics of the categories themselves.

ascended safely by Shakey). But as soon as Shakey could perceive the ramps' inability to stay still, they may need to be recategorized as non-ramps or subcategorized as not-useful-for-this-task ramps. To go with a cliché, in normal circumstances not everything is exactly as it appears to be, and Shakey would need to be able to both recognize objects for what they *appear* to be and be able to recognize them as something else if they didn't behave as anticipated. According to embodied theorists, this last part requires dynamic bodily interaction with the objects.

To explain and test this last notion, Beer culled a minimally embodied agent down to one particular categorical perception: the ability to distinguish a falling circle from a falling diamond when positioned underneath the falling object(s). Thus, his agent had only two categories to recognize (diamonds and circles). Additionally, it could only move horizontally and was given goals associated with each categorical perception. If it deemed the falling object to be a circle, the agent should try to center itself beneath the object so as to catch it; if the object were deemed a diamond, the agent should move out of the falling path to avoid it. And in these experiments the agent was roughly the same size as the circle or diamond, so from directly underneath it would be extremely difficult to determine the shape of the falling object. To achieve the desired goal, Beer designed the agent with an "eye" consisting of seven rays distributed across a limited area of the visual field (see figure 3.5.1). When the object crossed the path of a given ray or rays, the object was detected by a corresponding sensory "neuron." Simultaneously, these patterns of stimulation would be projected (Beer's terminology) to five interconnected interneurons, which would in turn project to two motor neurons determining whether the agent would move right, left, or stay still. [35]

[35] It is ironic that this test puts the AI back in the same sort of oversimplified, unrealistic environment in which Shakey was tested. But at least the AI had to deal with a more dynamic situation. Consider it a step back to go forward in a new direction.

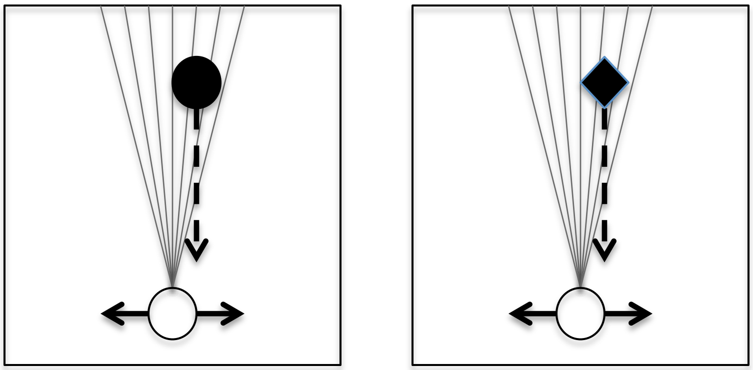

Figure 3.5.1 Basic setup for Beer's categorical perception experiments. The agent can move horizontally while objects fall from above. It uses an array of seven distance sensors to discriminate between circular (left) and diamond-shaped (right) objects, catching the former while avoiding the latter. (Beer, 2003: 213)

From these experiments, Beer discovered that the agent was successful only when it moved back and forth while attempting to decipher the shape of the falling object. If the agent stayed still it was either unable to determine the shape or able to do so only after the object was too close to avoid. In order for the agent to make a correct determination of the shape of the object and then take the appropriate catch-or-avoid movement, it had to oscillate horizontally beneath the object as soon as the object was detected, allowing it to break several lines of sight in different ways. In other words, it had to actively scan the environment. The reason why active scanning was crucial in detecting the shape, and why this is believed to be evidence of embodied cognition in action, has to do in large part with Beer's reliance on a steady-state horizontal velocity field (SSHVF) when programming the agent.

To understand the SSHVF, first consider how a circle or diamond would appear if it became frozen at a particular point during its fall. Should this happen, and given the agent's visual rays, the agent would receive constant input from the object that could be plotted on a particular (x, y) coordinate. If this occurred, the agent would move, at a constant velocity, from side to side in order to decipher the shape and, once the shape was known, it would move (again at a constant velocity) to one of its "equilibrium points"—either directly underneath the circle or to the side of the diamond. This scenario and reaction seems close to how Shakey interpreted its environment and competed its tasks. The objects were in fixed positions and Shakey could go about traversing the environment at a constant speed. However, since the circles and diamonds in Beer's experiments are not frozen, but falling, the agent has to do more than plot a single position of the object. It also has to determine the appropriate course of action as well as the speed at which that action should occur. Thus, when programming the steady-state velocity fields for two distinct falling objects with different associated actions, many possibilities and responses had to be included. This, in part, is the reason for giving the agent seven visual rays. When looking at figure 3.5.1, one can see how the circle and diamond break several lines of sight at their given positions. Move them up and the objects cross fewer rays; move them down and they cross more at once. With this detection capability, Beer and his associates developed an elaborate series of rules for the agent based on how it intersected with the lines of sight. Some of this had to do with detecting the shape (and whether to move toward or away from it), and some with the position of the object to determine the necessary speed to reach the equilibrium point.

They found that the SSHVF could not be utilized effectively without the agent actively scanning the object by moving back and forth beneath it at rates that had to vary based on the vertical position of the object as well as the input received from the visual rays regarding shape. For example, if the agent stayed still, the avoidance detection associated with diamonds would not occur in time for the agent to move out of the way, "the initial scan is delayed long enough that the trajectories become trapped." (Beer, 2003: 229) Whereas, when moving back and forth at variable speeds, the agent was able to settle into its equilibrium point more successfully by oscillating between avoidance and catch movements, and scanning multiple possibilities, as the object descended through the visual field—basically homing in on the appropriate position by seeing how the falling object was interpreted from several different angles and moving quicker or slower depending on the information received from the changing angles.

Because the agent's movement was determined by the continually changing position of the object, and the categorization of the object was determined by the continually changing position of the agent, Beer believes this to be an example of a coupled, dynamical system.

> As the agent's state evolves from its current point, the agent's resulting action and the environment's own dynamical evolution change the sensory input that the agent receives and modify the subsequent trajectories that are available to it. In this way, both the agent's dynamics and that of its environment continually shape the unfolding behavioral trajectory, as well as the future sensitivity of that trajectory to subsequent sensory input. (Beer, 2003: 236)

The environment shaping the behavioral trajectory and future sensitivity of trajectories relates to the idea of coupled feedback. The decision-making process is being continually shaped by the agent's interaction with its environment, and vice versa. The line between internal and external is not delineated in a nice, clear way where one obviously affects the other. Instead, the internal trajectories and representations (if they exist[36]) are derived from the perceived environment and the perception of the environment is shaped by the internal trajectories and representations. Both are continually entangled so that neither stands wholly separate from the other; both constantly affect and are affected by the other.

And relating back to his three primary components of situated activity in dynamical systems, one can see that this particular agent exhibits all three. 1) Concrete action—this one is a little easy to achieve in this particular experiment as the agent was designed to complete the very specific, concrete task or avoiding or catching the object. 2) Situatedness—the object and its path were the reason for the agent's movement, and constrained and compelled that movement depending on the categorization of the object. 3) Interactionism—the agent couldn't just scan the

[36] Some dynamical systems models deny internal representations.

environment (without movement) to determine how to behave; moment to moment, depending on how the object appeared, the agent would move quickly or slowly, or not at all, and how the object appeared depended on the movement of the agent. Interactionism is the most crucial component of the dynamical model because concrete action and situatedness are still achievable and important to the traditional model, whereas interactionism in this agent doesn't allow the internal to be considered in isolation from the external. What the agent is going to do depends on where it finds itself in the environment (directly beneath the object, slightly to the side, or farther to the side) and how it perceives its environment (falling circle versus falling diamond), and how the agent perceives the environment depends on what the agent is doing (moving quickly, slowly, or not at all). This aspect of continuous, interdependent relationship between environment, thought, and action is a coupled relationship for Beer. Figure 3.5.2 is a simple illustration of this idea.

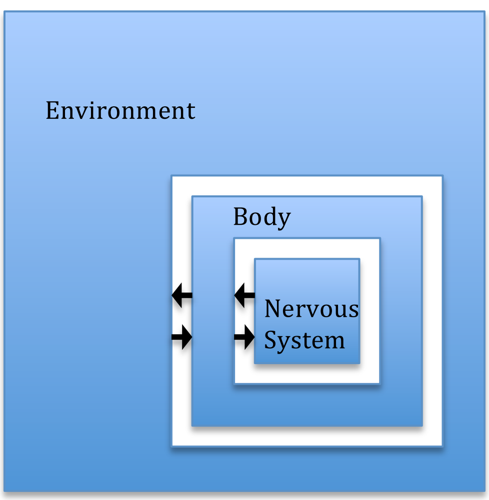

Figure 3.5.2 A dynamical perspective on a situated, embodied agent. The nervous system, body and environment of an agent are each conceptualized as dynamical systems that are in constant interaction. From this viewpoint, an agent's behavior arises from this interaction between these subsystems and cannot properly be attributed to any one component in isolation from the others. (Beer, 2003: 211)

The importance of this basic concept when put into action in designing an AI or when considering any cognitive model is that there is not a clear path of causality. Since the nervous system, body, and environment are continually interacting with and affecting one another about one another, the dynamical model Beer supports is a circle of causality with no distinctive beginning, middle, or end. Above this was described as "coupled feedback," but one can think of this as a feedback-feedforward causal relationship as well. This is quite different from the traditional model of Fodor which proposes a language of thought that exists independent of interaction with the environment and has a one-way causal route to establish the meanings of mental symbols (environment→body→nervous system/mind). And this is crucial to understanding why Brooks came to view Shakey as only a partial success and decided to abandon the sense→model→plan→act path of causality and cognition that dominated artificial intelligence designs based on the traditional model.

Due in part to work done by dynamicists such as Beer, as well as dramatic increases in computational capabilities, Brooks eventually replaced traditional models with what he termed a "subsumption architecture." Rather than the following the linear path from sensing, through modeling and planning, and finally onto action—so that there are several steps between sensing and behavior—the subsumption architecture connects them directly. To do this, the classical linear model is replaced with a layered one in which the control system has several distinctive actions available to it at all times, depending on sensor activity. This is different than the linear format of the classical model because there is not a singular path of processing that eventually leads to behavior, and the behavior is not reliant upon any sort of model or symbolic representation of the world at all. Instead, the subsumption architecture relies on an uninterrupted connection to the environment. Figure 3.5.3 shows the contrast between subsumption and traditional models, according to Brooks.

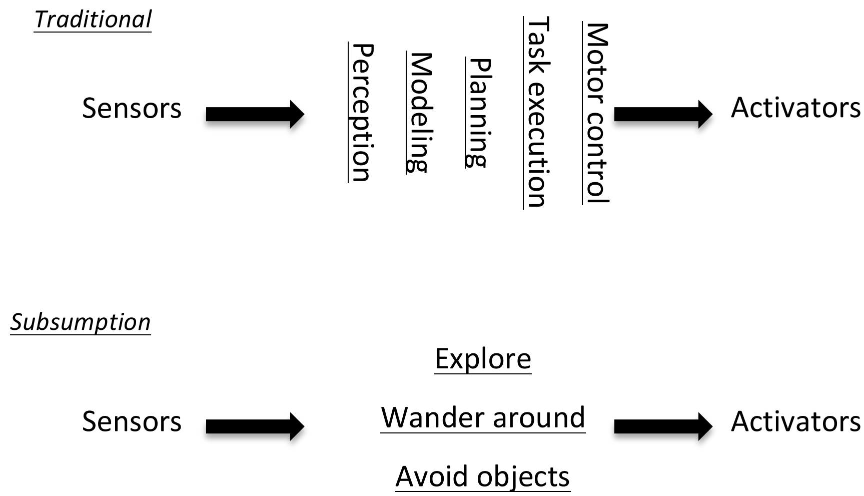

Figure 3.5.3 At top, a depiction of the sense-model-plan-act path of the traditional model. On the bottom, a depiction of Brooks' subsumption architecture. (Brooks, 1991: 122)

In the subsumption architecture there is not a one-to-one correspondence between input and action. The various layers have the ability to overrule, inhibit, or work in conjunction with one another. But there is a sort of hierarchy. The layers in the subsumption example in figure 3.5.3 refer to the behaviors of one particular robot Brooks constructed, Allen. Basically, Allen was designed to traverse an environment without running into things. But it was able to do this in a unique way due to its programming. Whereas Shakey may have required programming that allowed for the detection of three-dimensional shapes, modeling of the detected shape and correlating that model with a pre-programmed representations or models of certain objects, and then determining action based upon those models and previously encoded goals, Allen's perception-to-action path was very different. The layers in the architecture refer to the goals of the robot and they can only be accomplished by allowing the lower layers to be subsumed by the higher layers. Consider this in action.

Allen was a wheeled robot that had a circular base with twelve ultrasonic sonar sensors spaced evenly around it. Every second the sensors provided input connected directly to the explore/wander/avoid behavioral layers. If the sensors detected an object in the robot's path, the avoid layer would cause it to stop, then turn and begin moving forward again. At the same time, the wander layer would direct Allen in a new, random heading every 10 seconds. And the explore layer would detect distant objects and direct the robot towards them. But the key to the architecture was in how the layers could subsume one another. If the sensors detected an object in Allen's way, the avoid layer would inhibit the wander layer, preventing the robot from running into an object. Similarly, the explore layer could inhibit the wander layer in order to keep Allen on course to a distant object, and it could reorient the robot to remain on course if the explore layer had been interrupted by the avoid layer. As far as traversing the environment, Allen did so remarkably well and much more fluidly than Shakey. And from the observers' perspective the robot appeared to have a largely consistent, coherent pattern of behavior exploring, wandering, and avoiding. But, rather than a singular guiding "monolithic control" (Brooks, 1998: 54) as assumed under the means-end reasoning of Shakey and classical AI, Allen is more a "collection of competing behaviors." (Brooks, 1991: 144)

This last claim is crucial to understanding the value of subsumption architecture as a model for embodied dynamical systems. The idea of competing behaviors reflects the notion that a larger system, such as the human mind, is composed of smaller dynamical, continually changing systems that have their own goals and behaviors. In the case of Allen, each behavioral layer is its own dynamical system. When put together each layer still continues to have its own unique characteristics, yet, because each layer can affect and be affected by the others, the resulting overall behavior of Allen is not distinctly one layer or another, but rather a new, collective behavior.

The value of recognizing this trait is that it can allow for emergent behavior that cannot be fully predicted in the internal programming of the overall system(s) or agent. The behavior is emergent because it is not the direct algorithmic outcome of a single system, but rather the result of behaviors challenging one another based on the agent's involvement with and in the world. In terms of AIs based on Brooks' architecture, this has to do with the subsumption of one layer over another, bringing together the rules of competing systems and generating seemingly unpredictable behavior. Without seeing the environment in which Allen is placed and its trajectory within it, there is no hope of accurately predicting Allen's actions. And even observing the environment might not be enough as one could not be certain which objects Allen might choose to focus on when exploring. (Contrast this with Shakey's, where there was an ideal path it would follow in order to determine the success of the AI programming.) Plus, with Allen the unpredictability exists with only a few layers in the design. Should the architecture be expanded to 10 or 20 layers, the interaction and inhibition of behaviors grows exponentially, increasing the likelihood of highly unpredictable, emergent behaviors.

Whether referring to AI or human cognition, this is part of what makes a dynamical system different from traditional models. Michael Richardson and Anthony Chemero have delineated three key characteristics of dynamical systems, all of which can be seen in the relatively simple behavior of Allen (Richardson and Chemero, 2014: 39):

1. They consist of a number of interacting components or agents (homogeneous or heterogeneous).

   The physical form, locomotive capabilities, and layers of Allen interact with whatever objects in encounters in its environment.

2. They exhibit emergent behavior in that their collective behavior exhibits a coherent pattern that could not be predicted from the behavior of their components separately.
    Though the wander, avoid, and explore layers all have distinct goals and rules, Allen's exact behavior in a given environment cannot be fully predicted until it interacts with it and the layers begin to subsume one another. Yet, even without an exact mapping of behavior a coherent pattern will emerge in any environment as it avoids, wanders, and explores.

3. This emergent behavior is self-organized in that it does not result from a controlling component agent.
    Related to the last point, Allen's behavior will be based on the environment in which the robot is placed, coupled with its internal architecture. The goals will change depending on the objects in the environment as well as the emerging pattern(s) of behavior. The idea being that the behavioral control doesn't exist removed from worldly interaction (as assumed in classical AI and traditional cognitive science), but rather is based directly in it.

Brooks also echoes this last point from Richardson and Chemero. He refers to the notion of the top-down, removed-from-the-world control—such as a CPU receiving the input and generating a single model of the world that it then chooses to interact with in various ways—as "monolithic control." Brooks believes far too much evidence of rampant modality in human mental representations has been gathered over the last several decades (as discussed in section 3.3) and thus it is a fool's errand to construct an AI based on the assumptions of amodal, central processing removed from the sensory systems. Though he was guilty of doing this with Shakey, his view has changed considerably since those early days of AI construction.

> These [monolithic models] and other errors primarily derive from the naïve models based on subjective observation and introspection, and biases from common computational metaphors (mathematical logic, Von Neumann architectures, etc.)…a modern understanding of cognitive science and neuroscience refutes these assumptions. (Brooks, et al., 1998: 54)

At the heart of this critique by Brooks and the claims of other dynamicists is a rebuttal of the core principles that guided the construction of the GPS. This notion of a "monolithic CPU" complicates it a little since several in the traditional camp (notably Marr) propose a modular model of cognition that doesn't require or expect a single source wholly removed from the world. But the essential claim underlying the notion of a monolithic CPU is that cognition does not operate in an abstract, general-purpose rationality; it is grounded in physical interaction with the world that allows for novel goals and behaviors to emerge, and these goals and behaviors are continually shaped by the ongoing interaction the agent has with the world (and vice versa).

This division is also represented by two more recent robots: Honda's ASIMO and Boston Dynamic's BigDog. ASIMO—which stands for Advanced Step and Innovative MObility—is a humanoid robot built with the sense→model→plan→act architecture. The difference between ASIMO and Shakey is in the computational capabilities and physical construction. Compared to Shakey, ASIMO is capable of exponentially more computations at exponentially greater speeds and has much more elaborate joint mechanisms, sensors, gyroscopes, and motors that allow the robot to simulate human bipedal motion exceptionally well. ASIMO can jump, jump on one foot, climb stairs, jog, shake hands, and several other humanlike motions. And underneath it all, the robot is processing 3D snapshots of the world, relating the resulting representations to its symbolic programming, then generating motor commands and subsequent behavior from it. (Hirosi and Ogawa, 2007) This could seem like a win for the classical model and further evidence to the claim that the failure of robots like Shakey is not in the underlying model of cognition, but more in the limitations of existing technology. And, assuming more advanced technology will continue to develop with greater computational power, perhaps the classical architecture will eventually result in AIs that exhibit even more humanlike cognitive and behavioral abilities.

However, there are a couple of famous failures of ASIMO that may reveal a problem in the approach rather than the technology. At two separate technology shows ASIMO fell while attempting to climb and descend sets of stairs. At the respective events, the robot missed a step and fell backward when ascending and forward when descending. ASIMO is a humanoid robot with two arms and two legs, and walks upright, so one could assume this was a very humanlike loss of coordination. After all, people stumble and fall down stairs all the time. The failure was not in missing a step but in how the robot responded to the failed action. In both cases ASIMO did nothing to prevent its fall. The arms did not extend forward or backward to soften the impact, the leg(s) did not move quickly to reestablish balance. The robot appeared frozen in mid step as it fell and took the full force of the fall on its back and face, respectively. The reason for this was that ASIMO did not have an algorithm in its programming dictating how the robot should behave during a fall. Though it possessed gyroscopes that could detect the sudden imbalance and has limbs and joints that could be positioned to regain balance or lessen the impact, it did not have any programming matching such input with a desired action. And until a programmer includes an algorithm designed to deal with input corresponding to falling and the actions ASIMO should perform in response, the robot will continue to repeat similar failures in action. Dynamicists like Brooks and Beer could say this is why the classical model will never replicate humanlike behavior—because humanlike behavior is composed of coupled dynamical systems that are causally entangled and respond immediately and in concert to the shifting environment, as well as the shifting positions one takes in that environment.

Boston Dynamic's BigDog is an example of a much more elaborate dynamical, embodied model of behavior. The robot is intended to move similarly to a dog and looks like a mechanical version of a dog, minus the head and tail. Built along similar lines as the subsumption architecture of Allen, BigDog has no brain or CPU to speak of and nothing dictating its specific motion. Instead, the robot has many sensors continually tracking the state of its various components and limbs, and alters motion(s) as the situation demands in order to stay upright and moving. (Raibert, et al., 1998) The result is a robot that can handle sudden changes in situation much better than ASIMO. For example, in one experiment BigDog traversed a parking lot while taking a violent kick to the side of its midsection from a researcher. In response, the robot quickly splayed its legs to maintain balance—much as an organic quadruped would do—and continued on its path. Exposed to such a shove ASIMO would likely have tipped over. There is also a video of BigDog walking along an icy path. As its feet slide suddenly in different directions the robot immediately adjusts the directional force and placement of each leg to stay upright and continue moving forward. The movements are sporadic and sudden. Sometimes the robot ends up placing a knee joint on the ground in order to stay upright, and eventually it regains its full quadruped footing. Again, the action mimics the natural behavior of other four-legged creatures remarkably well, and the robot is doing so without any need of mental representations or the linear causal path of sense→model→plan→act. BigDog is just one example of dynamical systems robots designed by Boston Dynamics, and Brooks also has a more recent robot, Cog, that looks like a human torso and head and exhibits a great many humanlike movements (eyes that following moving objects, hands that work in concert, etc.).

Does this mean embodied theorists and dynamicists can rejoice in finally finding definitive empirical evidence against traditional cognitive science? No. One of the most glaring problems from the examples above is that they display an intelligence that falls far short of any standard for full humanlike cognition. More fluid movement? Yes. Ability to autonomously traverse and minimally interpret an environment? Yes again. And though these may be traits that are necessary for recognizing minimal cognition in another being (especially a nonhuman being) this is a very limited aspect of cognition. These robots and research do not exhibit traits of concept acquisition and comprehension, novel propositional attitudes, and more elaborate human locomotive abilities based on propositional attitudes such as holding an infant gently precisely because it is believed to be an infant. Certain observable behaviors are being simulated well, but much of what seems essential to humanlike cognition has yet to replicated or simulated sufficiently in AI—especially in dynamical systems designs. But, to be fair, Brooks, Beer, and others realize this.

Because embodiment theorists and dynamicists want to ground cognition in physical interaction with the world, their hope is that these robots represent the fundamental coupling between mind, body, and world that provides the basis for the higher forms of cognition just mentioned. That learning how an agent can interact fluidly and appropriately in novel situations may reveal greater insights into how one then acquires language as an infant, masters motor control, and then learns to apply concepts dynamically with the world—very much a ground-up rather than a top-down approach. There is a long way to go to achieve this, but they are confident it will be achieved someday. Just as with proponents of the classical model, dynamicists and embodied theorists believe efforts will be furthered significantly by future advances in technology and processing power.

The essential takeaways from this discussion are the failures and successes of these approaches when trying to construct a cognitive system. The traditional, or classical model immediately showed results in terms of replicating rational behavior, and continues to do so. However, its shortcomings have been revealed over and over again when attempting to create an agent capable of interacting with an ever-changing world. Assuming that real-world problem spaces are more unstable than the static logic problem space the GPS traversed and the situations in which ASIMO has been placed—as the factors, scenarios, and responses could change based on reasons outside the agent's control—an embodied artificial agent must display rationality and behavior that can cope quickly with rapidly changing environments. But, even if this last point is acceptable as a goal which all AI should strive to achieve, neither embodied, dynamical approaches nor more traditional models have yet to create an AI capable of exhibiting such flexibility in cognitive behaviors such as language acquisition, autonomous propositional attitudes, or the ability to reflect on mental states. Yet.

4. The metaphysical problem of mind-body-world dependence

So far the goal has been to provide the positive case for embodied cognition based on the many and variegated forms it has taken philosophically, and in terms of neuroscience research and artificial intelligence. However, it hasn't been one clear path, and this is part of the challenge facing embodied cognition—it's not always clear to proponents and critics what should fit under the "embodiment" umbrella. Nor is it clear exactly how and why supporters of embodied cognition consider it irreconcilably different than traditional cognitive science. The one constant has been the notion of coupling or coupled systems. But this particular term, and what it means, is perhaps the most ambiguous aspect of embodiment theories. In section 3.1 several versions of proposed "coupling" were discussed—Merleau-Ponty described perception and action as one coupled problem rather than two distinct problems to analyze, dynamicists describe the mathematical coupling of systems wherein there is no way to program or describe one aspect of a system in isolation from the other(s) nor is there causal priority, VTR claim color experience to be coupled with multiple sensory modalities, Johnson and Lakoff believe language to be coupled with sensorimotor experiences, and so on. Do they all mean the same thing when they refer to "coupling"? No. Gibson, for one, supported direct perception, which would link sensory stimuli with resulting concepts via little or no reliance on abstract inferences. On the other hand, some advocates of grounded cognition say thought still relies on a certain amount of amodal processing in perceptual concepts so long as it works in tandem with sensorimotor input. Similarly, the lack of causal priority demanded by dynamicists' models is not universally accepted, nor is VTR's denial of a pre-given world. With these discrepancies one should wonder if embodiment theorists are even arguing for a similar model of cognition at all.

Though there are clear differences in the way certain theorists use the term "coupling" there at least seems to be a similarity underlying them all. Specifically, that in order to provide an accurate model of cognition one cannot separate mental processes from body-world interaction. That is, body-world interaction should be thought of as an inseparable constituent of cognition.[37] At this point, such a claim is still too vague to separate embodied from traditional models because there is more than one way to interpret "constituent of cognition." And this is the critical point that needs to be examined more closely. The following chapter will detail how and why interpretations of constitutivity—specifically the metaphysical relationship between cognition and mind-body-world interaction—are the crucial feature delineating traditional cognitive science from embodiment theories as well as embodiment theories from one another.

The reason for playing up the notion of coupling in the previous paragraphs and chapters is because embodiment theorists use the term as a stand-in for "constituency." By linking mind, body, and world in the ways already examined, embodiment theorists are generally associated with the strong claim that this particular form of coupled mind-body-world interaction *constitutes* cognition. Generally, a constitutive relation can be thought of as a metaphysically necessary one. That is, if x constitutively depends on y, then it's metaphysically impossible to have x without y. For example, one might say that mental states constitutively depend on environmental factors. Thus, one can't have a CAT representation if one has never encountered cats. This is not a surprising claim, nor does it go against traditional accounts. The important

[37] Though to say "x is constitutive of y" is generally taken to mean x is part of what it is to be y, and y is nontrivially decomposable into parts, one of which is x, when embodiment theorists speak of constitutivity regarding mind-body-world coupling they seem to be saying that once x is decoupled from y, y ceases to be y any more. This will be discussed more as the chapter progresses.

difference arises at the type versus token level. Traditionalists would say the representation *type* constitutively depends on one's interaction with cats; but the representation *tokens* don't. One might hallucinate and see a cat, thereby representing CAT, even though there aren't any cats around. Or one might lose the ability to see, yet still retain a visual representation of cats. So mind-body-world interaction is metaphysically necessary regarding the semantic content of CAT, but not metaphysically necessary regarding mental representations qua mental representations (or one's ability to cognize at all). The token CAT still exists and functions in cognition regardless of an agent's current mind-body-world interaction. Embodiment theorists stereotypically want to make a stronger metaphysical claim. Specifically, that the ability to cognize at all, including tokens of concepts, is metaphysically dependent on mind-body-world interaction. This is because if mind-body-world can be shown to constitute cognition in the sense that continual interdependency between the three is metaphysically necessary it would undercut any traditional theory that calls for cognition to occur and be represented solely in an amodal network of formal symbols. In an amodal network, tokens of all concepts (even perceptually based symbols) still function without the need for the body-world constituents of the types—thought can, and often does, operate independently of mind-body-world interaction.

But disproving formal, amodal cognition is not easy to do, even with the presumed evidence discussed in the previous sections. This is because the sorts of metaphysically necessary coupled relationships generally associated with embodiment theories are exceptionally difficult to establish. Though there is no denying that the data collected in the behavioral and neuroimaging studies show relationships and correlations between modal processes and cognition, many supporters of traditional cognitive science believe the evidence only establishes increasingly more elaborate *causal* relationships between mind, body, and world. There is strong evidence of conceptual types having meaning due to embodied action, but far less evidence (if any) that tokens and cognitive capacities cannot function without it. And, should there only be causal semantic relationships at play, traditional cognitive models remain firmly intact because bodily states and perceptual systems don't become part of the mental state; they affect mental states and representations by providing semantic content but they aren't absolutely necessary for continued cognitive activity.

Another concern I'd like to introduce here, related to the type/token distinction has to do with the ever-present issue of brains in vats. Brains in vats are generally treated as a matter of permanent disembodiment. But there are at least two scenarios worth discussing regarding brains in vats: brains that have never been embodied, and brains that were once embodied but are now disembodied. This is related to the type/token distinction because even though all embodiment theorists will say a permanently disembodied brain cannot cognize—as conceptual types metaphysically depend on mind-body-world interaction—not all rule out the possibility that mental tokens (of embodied conceptual types) could continue to function in a brain that was once embodied but has now become disembodied. And if this is acceptable, the claim that cognition cannot exist without mind-body-world interaction, in all possible worlds, no longer holds.

With this in mind, this section is going to examine three main issues: 1) the metaphysical role of mind-body-world interaction in embodiment theories, 2) whether this metaphysical necessity is any stronger than traditional accounts, and 3) if embodiment theorists are consistently adopting the strong claim that cognitive activity is metaphysically dependent on mind-body-world interaction. Each of these will require plenty of examination and clarification since each requires an attempt to disambiguate a good deal of information. As will be discussed, I believe the majority of embodiment theories do not fully support the claim that all cognitive activity is metaphysically dependent on mind-body-world interaction. Because if a theory strives for nothing more than metaphysical dependency in terms of conceptual types, or permanent disembodiment, then it fails to achieve the differentiating goal of embodiment. The previous sections have been mostly charitable to the embodiment accounts. Now the burden of proof will become more evident and important for embodied accounts. This is because both sides in the debate admit the nomological role of mind-body-world interaction with the metaphysical caveat for semantic content. The expansion of metaphysical dependence between mind, body, and world for the very exercise of cognition is the actual point of disagreement, and thus the criterion in need of justification.

Questions of equivocation

Frederick Adams and Kenneth Aizawa—though they use different terminology—formalize the problem nicely in what they refer to as the "coupling-constitution fallacy." When they refer to "causal coupling" it is akin to a nomologically dependent relationship.

> The basic problem is that, in general, one cannot assume that a causal coupling with a process of type *Y* is sufficient to render the process coupling to *Y* itself a *Y* process. More specifically, we cannot assume that causally coupling a process *X* to a cognitive process *Y* is sufficient to make *X* a cognitive process. (Adams and Aizawa, 2008: 93)[38]

Adams and Aizawa are making the point that saying certain sensorimotor interactions with the world (governed by natural laws) cause the content of certain mental representations (such as the texture of cat fur) is not sufficient for establishing those sensorimotor interactions as *cognitive* processes. The mental representation of cat fur is cognitive, and though its content is nomologically, and possibly metaphysically, dependent on sensorimotor interaction with cat fur,

[38] Adams and Aizawa originally raise this issue against extended cognition theories, but the fallacy is just as applicable to the embodiment theories discussed here.

that dependency does not then make interactions with cat fur part of the cognitive process. In part, it's a disagreement over transitivity. The path is that interaction with the world (A) is needed to provide semantic content for mental representations (B) and that content is needed to allow B to be used in cognition (C). Embodiment theorists are saying that since A is needed for B and B is needed for C, A is needed for C. Adams and Aizawa are denying this because, though it may be true for semantic content about natural kinds, it's not true about cognitive activity involving that semantic content. Mental types and tokens help reveal this as tokens of CAT FUR, for example, could possibly continue to function independent of the sensorimotor interactions metaphysically and nomologically necessary for the type CAT FUR. Adams and Aizawa believe when many theorists use the term "coupling" they slip back and forth between natural-law causal explanations and metaphysically dependent explanations, confusing content with processes, and meaning with function. And in doing so, such theorists fail to recognize the importance of the distinction and muddy the waters for anyone trying to gain a clear sense of what a coupled relationship actually entails in terms of all-possible-worlds constituency.

To illustrate this point, they call out an example provided by Noë in favor of embodiment. In the example, Noë argues that the sensation of enjoying a sip of wine cannot occur without rolling the liquid across one's tongue. Because of this, "the liquid, the tongue, and the rolling action would be part of the physical substrate for the experience's occurrence." (Noë, 2004: 220) As Adams and Aizawa note, no delineation is being made between metaphysical and nomological dependence in this causal explanation, in terms of what is coupled to the representation WINE. Instead, "this last sentence [from Noë] appears to be a claim about constitution, one with a minor twist on the coupling condition, roughly, that there be some unique nomologically possible way of achieving something." (Adams and Aizawa, 2003: 94)

And this is what tends to happen in many uses of “coupling” by embodiment theorists: the coupling involves a causal chain leading to and from interaction with the external world, and it’s taken to be metaphysically necessary because the content of the representation comes from the sensory systems one happens to have, and such content could not have existed without them. With Noë’s example, there is no denying the role of the tongue and wine, and the interaction between the two in providing the “physical substrate” for the experience. But it’s not clear from armchair analysis if the physical substrate should be considered an inextricable part of the mental representation qua mental representation such that the concept WINE could not, in any possible world, be utilized without it. It seems crazy at this point to assume WINE could not operate in a cognitive agent who no longer has a tongue. However, to understand Noë’s claim, it will be necessary to consider the physical substrate in a different way. Specifically, he may not be willing to separate the neural processes and pathways necessary for generating the mental representation of the taste of wine from the overall sensorimotor system. Which would then mean the physical substrate also involves the brain and loss of those parts of the brain would entail a loss of the concept WINE, or properties of the concept, or its use in cognitive activities. This will be discussed more later in this chapter and the next.

In terms of cognition and one’s ability to be aware of having had the sensory and mental experience of wine (possibly including a mental representation of WINE, with all its sensory modalities), Noë is seemingly doing little more than providing a story of natural law dependency between the mental content and the physical state. In this particular example, he has little to say as to how or why one should consider the continued use of the token mental representation (of the experience) metaphysically entangled with the wine rolling over the tongue.[39] And this is important. If the cognitive awareness of the experience (complete with mental representations) just happens to have been caused by the physical entanglements between the brains and bodies we have, then it is not at odds with the traditional model. Conversely, if the mental representations of the experience absolutely cannot function apart from the physical mediums necessary for generating their semantic content, then he is making a claim greatly at odds with the traditional model. Perhaps this indistinctness is accidental, perhaps not. Either way, Adams and Aizawa are calling attention to an equivocation and ambiguity that pervades far too many uses of "coupling" in embodiment theories—an equivocation between nomological and metaphysical dependence. This needs to be addressed in order to understand precisely where Noë and other embodiment theorists stand in relation to traditional cognitive science because, though they may think they are directly opposed to traditional model, they may be saying nothing revolutionary at all.

The problem of equivocation is not one that can be sorted out cleanly, since there is nothing we can do about the terms others choose to use—such as "coupling"—and how they choose to interchange them. However, we can attempt to understand what they meant/mean by the terms they have already chosen. Therefore, the next part will look at the role of nomological and metaphysical dependence in several theories to determine if they are implying the possession of cognitive capacities derived from mind-body-world interaction entails nomological dependence between mind, body, and world—which does not rule out traditional accounts—or if they are making the more radical claim of metaphysical dependence such that these capacities cannot be exercised without continued mind-body-world interaction.

[39] I am saying "in this example" because Noë does provide more detail in other examples and works.

What are the different interpretations of constitutivity in embodiment accounts?

Since “coupling” is vague because it is not clear whether cognitive systems or just semantic contents are coupled metaphysically to body and world, let’s start by attempting to set forth a clear definition. I’ll start with a sort of stereotypical division based on how embodiment theorists present themselves and the traditional model. This will start by painting with a broad brush, but will get more meticulous as the discussion goes on. According to what I’ll refer to as the “weaker” interpretation, mind-body-world interaction is metaphysically coupled to content (and capacities based on that content) but not the ability for that content (or capacities) to function. In this case, the type CAT requires interaction with cats, but tokens of CAT exist and function just fine, even when the coupled interaction ceases. This is weak because it doesn’t rule out the possibility of cognition taking place without one or several of the semantically coupled parts, or completely different parts altogether (such as a once-embodied-now-disembodied brain). The stronger entails metaphysical coupling to content and function, or capacity and exercising that capacity. A brief analogy will help to illustrate the difference.

Consider a quartet. It needs four musicians to exist. If one of the musicians that constitute the quartet is not present, the quartet ceases to be. There may be three people that could potentially form a quartet once the fourth member arrives, but until s/he shows up, there is no possible way for the quartet to function as a quartet. In this way, the properties/characteristics that make up the quartet are metaphysically entangled with the natural object (or event) that operates as a quartet. It’s not simply a nomological fact that quartets as we know them happen to have four members; it is a necessary property than must be present in all possible worlds for the quartet to be present and play music. This point, or something like it, is held up by some embodiment theorists as akin to cognition and a point against the traditional model. The physical components as well as their unique functions *constitute* the quartet. Any attempt to remove or replace a member with a non-physical property/character/whatever, will result in the dissolution of the quartet's ability to work as a quartet. It is not just an issue of having things work together to create music—there must be four different instruments played at the same time for the quartet to operate. The physical presence is a necessary, defining trait of the quartet's continued existence in all possible worlds. The stronger interpretation of metaphysical dependence on mind-body-world coupling associated with embodiment theories can be taken to make a similar claim about cognition—remove body or world from the coupling and any concepts or cognitive capacities derived from the interaction between body and world cease to operate cognitively.

The weaker interpretation of metaphysical dependence in the coupled relationship would not deny the necessity of certain coupled interactions to create a quartet in any possible world. Namely, that there must be four musicians present at the same time, playing together, in order for a state of affairs to ever be designated as a quartet. But, what they would stress—and what the strong version often overlooks or denies—is that once the quartet has been physically instantiated and designated, it can be abstracted and still function. For example, let's assume all of the members of the initial quartet are human, and one of them comes down with a good ol' human stomach bug and can't make a performance. With that person absent, the quartet is absent. However, if a person in the audience were a musician of comparable talent, s/he might be able to fill in and the quartet would be present again. Or, instead of a person from the audience filling in, it could be a robot, an incredibly well-trained dog, a complete recording of the missing member's performance, or anything else that could fulfill the function(s) of a member of the quartet. In such instances, the quartet is abstracted because its existence is no longer based on the presence of the original physical tokens; once the type QUARTET has been established, it can be instantiated and work in any number of ways, so long as whatever physically constitutes it retains the same meaning, plays the same role, or fulfills the same function. And this is still assuming the concept of a four-piece musical group. That requires four concrete physical entities to couple together, but QUARTET can be abstracted further to mean the combination of four distinct things, with no clear connection to the four human musicians of a musical quartet. Life, liberty, property, and the pursuit of happiness are a quartet of concepts; Arthur Dent, Ford Prefect, Zaphod Beeblebrox, and Trillian, are a quartet of fictional characters. In these examples, the metaphysical dependence is on the relationship of four things, but what those four things are or could be, is contingent and often determined by nomological constraints (such as the physical traits necessary to play a four string violin). This is representative of the weaker interpretation because it differs from the strong interpretation by admitting metaphysical dependency only in terms of generating content or capacity, and denying that the continued application of the concept requires the original mind-body-world coupling.

With this in mind, one way to consider the differences in the interpretations of constituency is to think of the weaker version as focusing on a claim about cognitive capacities, around mental types, and the stronger version focusing on exercising those cognitive capacities, around mental types and functioning tokens of those types. If these interpretations are applied when attempting to explain cognition as it is normally observed in an embodied agent, it reveals a clearer separation between traditional and embodied accounts on mind-body-world interaction and its metaphysical role, as well as why the traditional account accepts the weaker version but not the stronger.

(Weak/Traditional) *If mind-body-world interaction provides content for mental symbols, then mind-body-world interaction is necessary for*

cognition.

(Strong/Embodiment) If mental symbols are derived from mind-body-world interaction, then every cognitive activity involving those symbols occurs only if there is mind-body-world interaction.

The traditional interpretation admits mind-body-world interaction can and does play an important role in cognition by providing semantic content for mental symbols, but doesn't say mind-body-world interaction *is* metaphysically necessary for continued use of those symbols.

Consider the mental representation CARDINAL. Its many properties may be derived partly from perceptual experience in that the association of RED and BIRD could have come from seeing a red bird and having someone else point at the bird and say "cardinal." In this way, the concept is directly derived mind-body-world interaction but the ability to retain and utilize the concept CARDINAL may not depend on continued access to or activation of the perceptual systems responsible for the initial input. Once translated to the amodal language of thought other cognitive processes can make use of the concept without continued mind-body-world interaction even though it was caused or based on embodied activity. Thus, sensorimotor activity can be metaphysically necessary for this cognitive content without cognition continually requiring it. Additionally, stimuli resulting in the experience of redness, birdness, and the audible sound of "cardinal," may not require interaction with any real birds, colors, or sounds. If once-embodied-now-disembodied brains in vats can be stimulated so as to simulate real-world interactions, or if the mental representations that constitute CARDINAL have been abstracted to the language of thought and can be activated by targeted neurostimulations, then the interactions necessary for the semantic content to persist and function do not require continued mind-body-world coupling. And, importantly, even if it were the case that RED, BIRD, and CARDINAL are metaphysically coupled to sensorimotor capabilities, and thus could not exist in a disembodied mind due to its inability to have continued interaction with the world, such a lack of sensory-based concepts does not entail a loss of overall cognitive ability without the added claim that all abstract thought is derived from embodied experience. Without that added claim, it might be possible that disembodied concepts derived from disembodied stimulations would result in concepts with very different semantic content, but there could still be minds, concepts, and cognition. Mind-body-world interaction has been reduced to mind alone or mind-world at most; there is no body involved. Having a body involved would affect the content of many representations and thus many propositional thoughts as well—it could be that scratching one's nose might not be a possible novel thought for a brain in a vat—but the fact that the brain is disembodied does not rule out the potential for cognition overall. I don't think most traditionalists would support this claim. I bring it up as an example of how failure to close the conceptual loop, so to speak, in terms of metaphysical dependence on mind-body-world interaction for mental types and capacities leaves room for various possibilities of disembodied cognition.

The strong embodiment interpretation requires that both the establishment of mental types and the continued use of subsequent tokens of mental representations require mind-body-world interaction. Initially, this may sound flawed because there are many instances of cognition that don't require continuous interaction with the environment, such as dreaming, meditation, hallucinations, or any other thought that occurs when one is immobile or has impaired or damaged sensory systems. But one way to construe the strong interpretation is that one must have and continuously maintain the *possibility* for mind-body-word interaction for cognitive capacities to function, in that one must maintain the neural capability to perceive colors, sounds, textures, etc., in order for those modal properties (and all concepts derived from them) to operate cognitively. This assumes modal-specific neural processing as part of a given sensory system.

Meaning that loss of one's eyes or damage to the optic nerve alone will not preclude mental representations with visual content caused by that sensory system before it was damaged. But if one should lose the modal-specific neural processing as well, then modal-specific concepts of that sort will not be possible in cognitive activity anymore. Another way to think of the importance of modal neural processes in this type of "coupled" relationship is to note how a patient who has lost an eye (or both) can have the optic nerve repaired, or an artificial eye attached, and could possibly see again so long as s/he retains the modal neural function. However, if s/he should have significant damage to the modal neural process, no stimuli (organic or artificial) will produce mental representations with visual content.

However, this is not yet enough to clearly differentiate the weaker from the stronger interpretation. Is the relationship established in the stronger version nomological or metaphysical? If one focuses solely on nomologically necessary properties, traditionalists might be willing endorse the stronger interpretation as stated. This is because if one assumes an innate language of thought, mind-body-world interaction and its role in the semantic content of mental symbols is a nomological problem. When considering the mechanical makeup of a cognitive system and the laws governing those mechanics, it's no surprise that one might lose the ability to generate and retain modal representations when the modal neural processes responsible for such representations are damaged. The physical systems that happen to be needed for one to visually perceive are gone, and this will have a lasting impact on one's cognitive ability to produce and retain mental representations with visual content. The same goes for all other mental representations containing modal semantic content. If one had not or could not utilize a given sensory system then the symbolic information derived from that system couldn't exist.

Nevertheless, this mechanical fact doesn't bind human cognition metaphysically to the human body. There is a metaphysical need for some mechanics to allow for mental representations to form (and hence any sort of conceptual and propositional thought), but which mechanics operating according to which physical laws is metaphysically arbitrary. If we imagine a car sitting on the side of the road out of gasoline, just because it lacks gas—a physical necessity that propels many of the other parts into action and allows the auto to become mobile—does not mean the automobile ceases to be an automobile when the tank is empty. The gas-less automobile is still an automobile system constructed in a very specific way, based on the properties of the materials that constitute it and physical laws governing the entire system. It is simply waiting for one constituent part to become involved in order to trigger a causal chain that will allow it to move. This is where the nomological versus metaphysical necessity becomes more evident: the importance of the gas, metaphysically, is in its causal role, not in its role as a liquid known as gas. It is metaphysically necessary that *something* cause the automobile to become mobile, but that something doesn't have to be gasoline. Electricity or hydrogen fuel cells, if used instead of gas, could play the same causal role of providing the energy needed to move parts of the systems in order to produce motion, yet both of those systems would rely on very different processes in order to do so. So the mechanical/nomological causal chains could be very different in each of the systems (since they rely on distinctive materials and chemical reactions), though the causal role in the overall system would be the same.

Prior to the development of electrical motors and fuel cells, an automobile designer may have believed the only possible way to make a machine mobile was with gasoline. Thus, s/he could have made the mistake of equivocating gasoline's apparent nomological necessity as metaphysically necessary in achieving automobile-ness. That is, the designer might have believed without a gasoline-powered engine one could not possibly create an automobile. But with the advent of more advanced propulsion technology, the designer would have seen gasoline could be removed from the automobile without stripping automobile-ness from the machine. The metaphysical necessity is that *something* causes propulsion. The nomological necessity then becomes using materials and forces that can be manipulated to achieve that propulsion. For a time, it was gasoline only. Now it includes other physical properties and processes, and may include many more in the future.

A proponent of traditional cognitive science could relate this to cognition a couple of important ways. First, with the neural structures we happen to have, in the bodies we happen to have, and in the physical world in which we happen to live, the most obvious and ubiquitous source of semantic content is mind-body-world interaction. If one tokens CAT or DOG or HAT it is because at some point a cat, dog, or hat was perceived in some way, somewhere in the world. But, the traditionalist could say, it doesn't necessarily have to be this way. This relationship between perceptions and mental representations is essential, but only in regards to the concepts generally derived from perceptions. Should we uncover ways to stimulate neural networks in causally similar ways to perceptual systems, a brain could receive input without any need for the body. The concepts would not have similar sensual properties as those of CAT, DOG, or HAT, but the lack of sensory input as we now know it, does not rule out the possibility of concepts caused by the type of "sensory" input a disembodied brain might receive. I have no good way to describe such concepts or properties, but it might be possible that disembodied concepts could exist based on whatever stimuli the disembodied brain receives. All that matters is that *something* plays the causal role of providing the input needed for the system to generate semantic content. (And, as mentioned before, even if the tokens could not relate to natural kinds—if, say, no brain in a vat could ever token NOSE or SCRATCH—presumably it could still token other concepts based on the input it is receiving.) Whether the causal input involves organic optical nerves, laser relays, digital processors, or anything else imaginable does not matter. This, again, is why brain-in-a-vat scenarios are possible under the traditional model. The causal relationships, qua causal relationships, are the metaphysical necessities, but how those relationships come to fruition are the nomological necessities that can vary from cognitive system to cognitive system.

Secondly, even if the traditionalist holds that certain mind-brain-world causal interactions are metaphysically necessary for certain conceptual types having the content they have and certain representations having the capacities they have—why DOG refers to man's best friend or HAT can be combined with different material and shape properties—it doesn't follow that any particular tokens of one of these concepts or mental states constitutively depend on the causal transaction. Similar to how a photograph of a glacier requires a photographer and her camera interact with the actual glacier, reproductions of that photograph—whether on contact sheets, paper printouts, or digital versions—exist and function separate from the glacier or the physical processes and mechanisms necessary to take and print the original photograph. This is not to say concepts are images, but the point that tokens of types do not need to be constituted in the same way in order to retain the same content or operation. There may always be some sort of causal interaction needed to create the tokening of a conceptual type, but this dependency is merely contingent.

Due to this, if embodiment theories are offering a unique, non-traditional account of cognition, then they must endorse mind-body-world interaction as a metaphysical necessity in terms of the exercise of conceptual tokens and cognitive capacities, and not just conceptual types. The key then becomes the middle part of this coupled interaction: the body. There must be a mind[40], with some ability to process information, and information may only be acquired via some sort of input. Remove either and cognitive function is impossible. Even if the input is minimal and only provided for a short amount of time before being permanently disconnected or disabled, with no input of any kind, ever, there would be nothing to think about; there would be no semantic content for any mental state types or tokens.[41] So the differentiating metaphysical necessity is the body. I realize this sounds like a redundant point since the theories are all about embodiment, but they haven't been clear about this metaphysical necessity, nor have they been clear about what entails a "body." So now it's time to discuss what makes a body a body.

A stubborn supporter of embodiment might look at brain-in-a-vat scenarios and say that electrical stimulation of specific parts of the brain in order to simulate sensorimotor processes or generate any semantic content is still "embodied" activity. A much different form of embodied activity, but still embodied because the instrument and medium used to provide the stimulation are acting as a de facto sensory system and thus as a de facto (though minimal) body. If an embodiment theorist should make such a claim, I believe s/he would be wrong and using far too broad a criterion for "body." The type of body embodiment theorists generally mean when discussing mind-body-world interaction is not just a mechanism that provides input to the brain. The body they require is a *subject's* body, which has several criteria. Firstly, a subject's body is his or hers alone, providing input to and output from only that subject's mind. An external implement used to stimulate the brain could potentially provide identical input to multiple minds at the same time, thus violating this criterion. And if this criterion is not in place, and one wants

---

[40] I am saying "mind" here and not "brain" because one embodiment approach may be to consider the brain as part of the body once its functions and abilities have been delineated due to bodily interaction with the world. This will be discussed briefly later in this chapter and in more detail in the next.

[41] This is not presuming wholesale empiricism. There may still be innate cognitive abilities, but no content.

to claim the implement is a de facto body, then the line between body and world becomes exceptionally difficult to separate. If the mind has no ability to control the implement and receives the exact same information from it as any other mind would, then, from the perspective of the mind, the implement operates just the same as any other natural object because the mind can do nothing but passively receive the information from it. Thus, the implement is akin to the external world—not an agent's body—and once again there is only mind-world interaction taking place.

Secondly, the body must be under the control of the subject. All the talk of sensorimotor affordances loses its value if not based on the underlying assumption that the subject is capable of choosing how to maneuver through the world. The same goes for many of the concepts Lakoff and Johnson bring up. LEFT, RIGHT, FORWARD, BACK—all of these (and the many concepts based on them) are only possible if one has a body with contrasting sides, where left can be moved independently of right and where one has the possibility to turn and see what was previously behind oneself. Not only are these bodily concepts, they are only attainable through one becoming aware of one's body in relation to its maneuverability and relation to other objects in the world. In short, proprioception plays a vital role and it is not possible unless one has control of one's body.[42]

So, with these clarifications of the stronger interpretation in mind, there is now a clearer distinction between the weak (traditional) and strong (embodiment) interpretations of constitutivity. To restate, the two versions are:

> *(Weak) If mind-body-world interaction provides content for mental symbols, then mind-body-world interaction is necessary for cognition.*

---

[42] There are impairments wherein a subject may lose proprioception yet still have control of his/her body, but I do not know of a case where the inverse is true. If a subject has proprioception, s/he must have bodily control.

(Strong) If mental symbols are derived from mind-body-world interaction, then every cognitive activity involving those symbols occurs only if there is mind-body-world interaction.

The weak interpretation, as used by traditional cognitive scientists, allows for mind-body-world interaction to be metaphysically necessary for cognition, but primarily in its role of acquiring and providing semantic content for mental types and the cognitive capacities resulting from those types. The strong interpretation, as used by embodiment theorists, requires that mind-body-world interaction be metaphysically necessary for mental types *and* the resulting tokens, for both content *and* the exercise of that content. Loss of body entails a loss of the ability to exercise cognition (and one's body must be a self-controlled vehicle for interaction with the world that provides input uniquely to the mind entangled with it).

This delineation between strong and weak interpretations helps categorize theories more explicitly that are otherwise on the fence in terms of the metaphysical relationship in mind-body-world coupling. As will be discussed, there is a spectrum of embodiment stances, many of which do not meet the criteria for the stronger version. Failure to meet the stronger version doesn't mean a theory isn't providing a valuable counterpoint to most traditional theories, but it does mean that theory is not as iconoclastic as presumed.

Examining metaphysical dependence in embodiment theories

Since there is no necessary order in which to examine the interpretations of the metaphysical role of mind-body-world interaction, I am going to roughly follow the order in which the theories have been presented in this work. And I will only look at the most influential accounts rather than every philosopher or perspective mentioned. The goal will be to determine whether a given approach subscribes to the weak or strong interpretation to provide a filter of sorts that will help to determine which theories are really at odds with the traditional model and which ones are not. However, in pursuit of this goal, it will become obvious that not all theories fit cleanly into the strong or weak group. At minimum, the weak group has many variations dependent on the causal role bodily interaction with the world plays in cognition. Some will say connections to and interactions with the world must occur for conceptual content. A variation on this will say somatic concepts derived from such interactions are necessary for all abstract concepts. And, and the same time, everyone is willing to say the brain itself—as an essential constituent of cognition—cannot or would not develop properly without mind-body-world interaction. These stances are not at odds with traditional cognitive science. Every variation of the strong stance is clearly opposed to the traditional model, but we'll also see that the more radical embodied theories endorsing the strong view are endorsing a sort of relationalism. Let's begin with Gibson.

Gibson

Gibson's approach is a great place to start because of his incredible influence on this debate. The best place to begin looking at his theory is his insistence on direct perception. According to his ecological theory of perception, the contents of perception are not underdetermined and thus do not require inferences derived from a language of thought (housed and exercised in a central processing system) in order to become contents of thought. And, since inference *is not* essential to perception, sensorimotor interaction with the world and the subsequent changes it creates in one's ambient optic array *are* essential. Shapes, sizes, textures, distances, motion—all of these require movement of some kind from the observer in order to be perceived somewhat accurately. Without movement of some kind revealing the variances and invariances in one's environment—be it full-body motion or subtle eye movements—perception would be ineffectual for Gibson in generating mental representations.

Consider the example of the desk and the objects on and near it used to illustrate Gibson's theory in chapter 3. Assuming his theory to be accurate, one's visual awareness of the desk and objects, and all of their spatial properties, results from the motion of standing, sitting, and moving while taking in the apparent changes in the shapes, sizes, and locations of the objects in relation to one another, and one's own vantage point. Simply stimulating a brain-in-a-vat cannot, in any possible world, generate such mental representations. Additionally, when affordances are brought into consideration, concepts have the semantic properties they have only in lieu of one's potential bodily interaction with the objects to which they refer. Thus, BOOK has properties of readability based on one's visual abilities, and turnable pages and overall graspability based on one having a certain hand shape and dexterity. Remove the bodily interaction and the concept is either gone or cognitively useless. And since direct perception denies the role of inferences from an innate syntax of thought, there is no possibility of deriving a similar concept from other non-embodied information. Gibson might allow that if one had radically different sensorimotor abilities one would have radically different concepts. However, with no sensorimotor interaction with the world, no such concepts could possibly exist. And without the myriad concepts derived from sensorimotor affordances, no inferences or cognitive function whatsoever. Thus, he seems to be making the strong metaphysical claim that cognition occurs only if mind-body-world interaction occurs.

But this is where it's important to remember that a traditionalist theory like LOT is willing to accept that mental representations of GRASP and TURN (that can become properties of BOOK) require interaction(s) with the world to have the semantic content they have. Though Fodor disagrees with direct perception, he doesn't deny the metaphysical role of sensorimotor activity in allowing certain concepts to exist. This means a traditionalist wouldn't deny the last claims from the paragraph above—without sensorimotor interaction, no such concepts would be possible and concepts would be different, and if one's body were significantly different then it could certainly affect conceptual capacities. Though Gibson's direct perception provides a very different account of concept acquisition compared to someone like Fodor, there is ambiguity in Gibson's theory that should be challenged to see how different the theories are regarding concept retention and function. Namely, Gibson's theory does not rule out the possibility of cognition continuing without the possibility of mind-body-world interaction once affordances have been directly perceived. Even though embodiment is metaphysically necessary to provide proper brain development and cognitive content, it is unclear if he believes the continued exercise of those cognitive capacities to be metaphysically dependent on continued sensorimotor interaction. Brains in vats could have cognition, so long as they have something to cognize—so long as they were once in a body that interacted with the world and allowed the individual to derive affordances from his/her/its experiences, it might still be capable of thinking even when disembodied. Thus, Gibson, though making radical claims in terms of perception and mental content, cannot be said to endorse the strong interpretation. Because his theory was early in this debate, and he was not responding to philosophers like Fodor, I'll move on from Gibson. But it's useful to introduce this problem of ambiguity with him, because nearly all who have adopted his theory, or built on it, have failed to address it in the decades since his work.

Valera, Thompson, and Rosch (VTR)

Since VTR are largely to thank for proliferation of the term "coupling" in embodied theories, finding out where they land on the metaphysical role of coupling in cognition is certainly worthwhile. The first thing to recognize is that, contrary to Gibson, VTR deny direct perception. Gibson held a sort of separation between mind-body and world. For him, though an agent must interact with the world to acquire information from it and derive any resulting mental representations, the world is taken to exist independent of mind and body. In VTR terminology, Gibson views the world as "pregiven" in that it requires no interaction or justification for it to be as it is—it is the objective environment into which our subjective minds and bodies are thrust. VTR do not accept the world as pregiven and instead put greater emphasis than Gibson on the last hyphen in mind-body-world coupling, saying the world is brought forth via this coupled relationship. Their paradigm example of color experience is supposed to reveal this. As they said, "Perception and action, sensorium and motorium, are linked together as successively emergent and mutually selecting patterns…our colored world is brought forth by complex processes of structural coupling." (VTR, 1991: 163-164) This provides a clear point of difference between Gibson and VTR, but they still face the same challenges and ambiguity of Gibson's theory. VTR's account is almost entirely focused on conceptualizations of, and derived from, the world. And they do not directly address fundamental metaphysical or nomological requirements for cognition. But this doesn't mean clues aren't sprinkled throughout their efforts.

Unlike Gibson, for VTR it makes no sense to discuss the "the world" as if it is out there offering myriad affordances available in the same way to all people. With the notion of a world "brought forth," VTR are throwing out any attempt at trying to imagine the world as an objective source of information at all. Instead, the information received from the world is linked so tightly to sensorimotor activity that it can never reveal wholly objective facts beyond the scope of one's sensorimotor capabilities and interactions. This is not to say there are not agreed upon facts revealed about the world—such as the circumference of a tree stump or its height—but these facts are epistemically and metaphysically derived from an individual (or individuals) seeing it, touching it, and interacting with it in various ways. Thus, at best we achieve consensus about aspects of the world, but consensus (in VTR's model) is nothing more than individuals finding common ground among one another's subjective experiences.

It follows from this that one never encounters the world *as such*. One always encounters the world through the lens of one's sensorimotor capacities, and these capacities are inherently limited and biased due to one's physical form and abilities. Consequently, our engagement with the world both reveals and shapes it because our perceptions of the world are derived from our embodied sensorimotor systems, and are affected by our past experiences. For example, only objects of a certain size and shape are perceived as graspable because of the size, strength, and dexterity of one's hands. This, they say, is why there is no one-to-one correspondence between percepts and natural objects. Several individuals may encounter the same apple in a field, but depending on each person's physical abilities, sensory system processes, and history, each may perceive the apple differently in terms of color, graspability, edibility, etc. This is also why the green pepper can appear green even when reflecting different waves of light. The perception is not based solely on the data received. The data is vital, of course, but just as it affects sensorimotor systems and mental representations it too is affected by sensorimotor abilities and other mental representations.

With these characteristics in mind, VTR are claiming mind-body-world interaction plays a unique role in cognition compared to traditional theories, altering the causal path for metal representations as well as the independence of the external world. But they are not clearly endorsing the strong view. The strong view requires continued mind-body-world interaction in order to retain any cognitive act or ability utilizing symbols derived from mind-body-world interaction. But VTR are not making this claim. They are endorsing a weaker conditional statement. For example, VTR claim meaning is derived solely from the external input *and* the body—meaning that if a brain-in-a-vat were to receive unfiltered input directly from the world, without the body acting as the intermediary, the stimuli would have no meaning from which semantic content could be derived. This is really only a conditional statement in line with the weaker interpretation. *If there were not mind-body-world interaction, then an agent would not have mental content*. The stronger interpretation is not met or needed. They are basically saying that if a disembodied brain were grown in a lab and stimulated, it would be stimuli with no content; input without information. But, just as with Gibson, this does not rule out the possibility of a once embodied brain retaining cognition when disembodied. And if that is a possibility, there is no metaphysical dependence between cognitive *function* and mind-body-world interaction. Content, yes; but not continued use of that content.

In addition to failing to meet the strong interpretation, VTR's denial of a "pregiven" world brings up another challenge. Denial of a pregiven world means the denial of objective facts existing independent of the perceiver(s). The problem is that even if what the world means to a particular agent depends on the agent's embodiment, that doesn't necessarily rule out the existence of external properties or objects. If it does, and there is no objectively deducible world independent of perceivers, then VTR are leaving themselves open to the same criticism Locke faced from Berkeley centuries before, and they could be taken as endorsing a sort of idealism—and this might create a significant problem. Namely, VTR and (as far as I know) all embodiment theorists want to say cognition is born out of bodily engagement with *the material world*. And since the world is a metaphysical necessity in addition to the body (if one is trying to provide an embodied theory that is truly differentiated from traditional cognitive science), removing the external material world from the equation would prohibit cognition from taking place. Even though one might be able to make an elaborate case for an idealist theory of embodiment, this likely isn't what they want to say—and they have not done an effective job of ruling it out. Being charitable, one could take them as saying the world has no meaning for an agent without the coupled triad. This way, perhaps they avoid committing themselves to any claims about the supposedly objective properties of the world. It's not that there aren't real objects, just that they only become meaningful to us via embodied interaction. So, to speak of them in a disembodied, objective fashion is literally nonsensical. But this isn't particularly satisfying either.

Though I believe VTR are unintentionally endorsing a weak interpretation of the metaphysical dependence between cognition and mind-body-world coupling, I don't think their original intent was to tackle the metaphysical consequences of embodied cognition—and this is largely responsible for the ambiguity in their metaphysical claims. In the introduction to *The Embodied Mind* they explicitly say, "We do not intend to build some grand, unified theory, either scientific or philosophical, of the mind-body relation." (VTR, 1991: xviii) Theirs is an account of experience aimed at explaining conceptualization and mental phenomena, while deferring the heavy metaphysical lifting to others. I do not state this an excuse for them, but rather as a point of clarification and frustration. In order to clearly differentiate their theory from traditional cognitive science, a strong metaphysical claim about the relationship between mind-body-world interaction and cognitive activity must be made (or at least attempted). Overtly, they did neither. And then several embodiment theorists picked up their "coupled" torch without realizing that much still needed to be answered.

Glenberg

When considering Glenberg's indexical hypothesis on how linguistic symbols acquire meaning, I believe he too fails to meet the requirements of the strong interpretation. Just as a reminder, the indexical hypothesis claims linguistic symbols become meaningful through a process in which words are mapped to "perceptual symbols" based on affordances derived from those perceptual symbols. He is appropriating Gibson's notions of affordances, but his addition of perceptual symbols provides novelty to his theory.

According to Glenberg, words and phrases are not mapped to an amodal system of thought that allows one to infer and derive meanings of words based on its own underlying syntax. Instead, he takes the language of thought (if there is to be such a thing) to be distinctly modal and analogical. The linguistic symbols one can have are never represented apart from the perceptual system(s) responsible for their creation, and thus are never abstracted away completely from the original perceptual processes and neural activations. They are not separable because they become intimately linked with perceptual processes in the following way:

Stage 1: Words and phrases are indexed (or mapped) to perceptual symbols.

Stage 2: Affordances are derived from the perceptual symbols.

Stage 3: Linguistic symbols become linked (inextricably) with affordances, thus yielding an understanding of the word or phrase.

Once the affordances and symbols are linked and understanding occurs, sounds and shapes of language become more than mere sounds and shapes. Now they have intentionality—they are about something; they *mean* something because the symbols are linked to specific perceptual experiences and underlying neural states. So the meanings of objects, events, and sentences comes from what a person has done and can do with them, physically, rather than being derived only from mental symbols and the supposedly amodal rules governing them. Without this physical grounding and meshing, the symbols would be nonsense.

Of vital importance to the indexical hypothesis is its opposition to the idea that language is mapped onto the world. One could argue that an individual starts with a definition of a given term or concept and then finds its referent in the world. For example one could start with a definition of “stool” as “something sturdy to stand upon” and could then explore the world to find such an object. So the affordance is determined because of the definition, not the other way around. But Glenberg claims this is not acceptable because the terms of definition acquire their meaning(s) based on one’s ability to extract affordances from various sensorimotor systems and experiences. The linking/meshing that occurs allows users of a language to discriminate between appropriate and inappropriate uses of terms and phrases within the language based on physical potentials derived from experience and not just definitional and inferential rules applied to experience. Thus, if one began with language and syntax alone, the definition of stool would be useless because at least the notion of “standing upon” would lack any meaning to a person who had not derived the affordance of standing on something via some sensorimotor system. This is why the indexical hypothesis is analogical—the symbols utilized in cognition are never extracted away to a fully amodal form, as they are with digital translations of music. Analogically, the symbols remain inseparably tied to the sensorimotor systems to the point that when one utilizes a given linguistic term that has been meshed with a perceptual symbol in cognition it is only possible due the reactivation of parts of the initial sensorimotor system from which it was derived (which has some empirical backing based on several of the experiments covered in chapter 3).

If the meshing relationship were merely a nomological necessity, one would have to allow for the possibility that this relationship could be replaced or replicated in a different possible world. But Glenberg denies this possibility. The analog nature of the indexical hypothesis entails the physical make-up of the symbols—including everything from the embodied form, physical potentials, perceptual systems, neural processes, etc.—constitutes specific semantic content as well as overall cognitive function. He is not only saying that one's language would have different meanings if meshed, say, with only electrical impulses to a disembodied brain. He is saying a disembodied brain, if stimulated, would result in only meaningless neural stimulations; language can only function in an embodied being because its meaning and function is continually derived from sensorimotor interactions with the world. It is not just the way language happens to be; it is the way language must be.

Yet again, this is a solid case against cognition in a permanently disembodied brain, but not in a once embodied brain. Here is another embodiment theory attempting to separate itself from traditional cognitive science based on a metaphysical claim about mind-body-world interaction in providing mental content. But, once more, it's not clear that a mind—operating analogically as described above—could not retain that function and the content therein, even when disembodied. And if the body can be removed and cognitive function and faculties still remain, cognition itself is not metaphysically dependent on embodiment. Cognitive content, yes; cognitive function, not necessarily.

Noë and Hurley

The enactive model developed by Noë, and Hurley echoes VTR by agreeing with Gibson's claim that action is essential for perception, and also denying direct perception. Similar to VTR, they say there is no one-to-one correspondence between input and experience or perception, and affordances do not exist independently, out in the world. Rather, the very existence of affordances depends on embodied interaction with the world. Thus, any brain-in-a-vat scenario in which the brain is imagined to have a visual experience similar to an embodied brain is impossible. Neural activity may be necessary for perception and visual experience, but neural activity alone is not enough to produce visual experience of anything. It is not sufficient because vision and other perceptual systems and experiences are intimately linked with the body in that a given environment is perceived in a certain way due to one's potential embodied interaction with it. In order to provide any unique semantic content, the thing or things being perceived must be interpreted, differentiated, and categorized by the perceiver via the perceiver's unique perceptual modalities. According to Noë and Hurley, these facets of perception are derived from one's potential interactions with the object(s) or, as they say, one's "sensorimotor contingencies."

This is the primary reason they disagree with Gibson's notion of direct perception. Rather than placing affordances somewhat objectively in the world, they say sensorimotor contingencies couple affordances to the body because the contingencies are based on the unique body one has. This then blurs the distinctions between mind-body-world because "the substrates of consciousness [and cognition as a whole]—in particular of visual perceptual consciousness—seem to cut across the brain-body-world divisions." (Noë & Thompson, 2004: 25-26) No body, no sensorimotor contingencies. No sensorimotor contingencies, no way to perceive the world. No way to perceive the world, (presumably) no way to exercise cognition.

Alone, if sensorimotor contingencies play the fundamental role claimed above this still would not be enough to associate Noë and Hurley with the strong interpretation. If semantic information were only derived from sensorimotor input, it might be that stimulating a brain in a vat in a similar way could result in similar mental representations. However, the claim is that the body one has and how that body can interact with the world shapes the input itself. The type of body one has might be able to change dramatically, and with it some nomological properties regarding sensorimotor contingencies, but one must have a body to have any mental representations. Thus, embodied interaction with the world is a metaphysical necessity in terms of generating metal content, but it's not yet clear that a once embodied brain could not still cognize once disembodied. This is very much in-line with the trappings of VTR and Gibson. But let's look deeper. Some aspects of the strong interpretation may come from Hurley's notion of horizontal modularity.

Fodor's vertical modularity claims that each module performs a function, generating or affecting a mental representation, and then moves the representation on to the next module, creating a unified representation composed of properties derived from these linear, independently operating modules that eventually resides amodally in the brain. Hurley's horizontal modularity, on the other hand, envisions the modules as overlapping and mental representations existing in and among the layers. As Hurley was quoted in section 3.2,

> "Rationality reconceived in horizontally modular terms…does not depend only on internal procedures that mediate between the input and output…it depends on complex relationships between dedicated, world-involving layers that monitor and respond to specific aspects of the natural and social environment and of the neural network, and register feedback from responses…On this view, rationality is a higher-order property of complex patterns of response, which emerges from the layers of direct dynamic couplings between organisms and their structured environments." (Hurley, 2001: 10)

The idea is that cognition is a network of input-output-input loops that continuously rely on external and internal feedback. There is no abstraction from the perceptual modules in terms of processes—wherein a part of the brain can utilize perceptually derived concepts wholly removed from the perceptual systems that provide(ed) the semantic content. Similar to sensorimotor contingencies, the result of horizontal modularity is perception and action becoming inseparable because mental representations exist as connections/relationships among the layers. Hurley described mental content in this model as "leaky" because it never exists in one module; it always leaks into others and is never abstracted to an amodal language of thought. And since the modules overlap, rather than linking linearly, removal of a sensorimotor module would have reverberations among all other modules and mental content based in the relationships between those cognitive layers. Removal of all perceptual modules would also then entail a loss of all perceptual mental states—and not just a temporary loss of that type of mental content or temporary loss of its use, but also an inability to ever possess it again. Hurley is saying there is no semantic content possible without embodied engagement with the world.

But is this strong or weak? When looking at this to determine the strong or weak interpretation of the metaphysical role of mind-body-world coupling, the issue is in how one considers "sensorimotor modules." If the modules are entirely brain based, then we end up back in the same cycle as discussed with Gibson and VTR—the module will not acquire any content without an agent's interaction with the world because the overlapping layers of the module(s) only generate mental content when they are coupled with body-world interaction. But generation is different from retention, and it's not clear that neural sensorimotor modules could not retain content once disembodied or disconnected from somatic sensorimotor systems. Even if we throw out the possibility of abstract mental representations, as Hurley wants to do by not allowing any abstraction from the modules to other areas or symbol systems of the brain, it is possible to imagine a scenario in which the sensorimotor modules of a developed, once embodied brain could retain their function and analog content even when disembodied. And if so, the activation of multiple areas/layers identical to the activations that occurred in embodied sensorimotor activity should allow for sensorimotor-based mental content to be exercised in cognition. The only way to rule out such a possibility is to say that sensorimotor modules must include the neural *and* somatic parts of the system. If so, then this would certainly be an endorsement of the stronger interpretation and no cognitive activity would be achievable, in any possible world, without mind-body-world interaction. However, I can't confidently say she is doing this. There is not an instance where she asks readers to consider the neural brain-based modules as body-based modules once activity has been horizontally mapped across the brain. Though she admits modules are "leaky" and affect one another as well as how one perceives the world, that is not enough to indivisibly link neural and somatic systems. When compared to VTR and Gibson, Hurley is making a stronger claim, but not (definitively) the strong claim. Noë, however, might be supporting the strong interpretation.

I have grouped Noë and Hurley together because they have published together and endorsed the "enactive" model. But it's worthwhile to consider Noë a little separately here. One of his favorite analogies for cognition is dancing. Just as dancing does not reside in the muscles of the dancer removed from the act, cognition does not reside in neurons removed from action. A dancer is only a dancer when dancing—when moving about the world in a distinct way, guided by his/her intentions and motor capacities, constantly adjusting based on the his/her position in, and interaction with, the world. He describes cognition similarly—as an activity. Cognition is something one does; it's a relationship between an agent's sensorimotor capacities and the world. He very clearly says consciousness is not in the head. "A neuroscience of perceptual consciousness must be an enactive neuroscience—that is, a neuroscience of embodied activity, rather than a neuroscience of brain activity." (Noë, 2004: 227) In line with this, as mentioned before, he claims that perceptual experiences constitutively rely on sensorimotor activity to the point that brain states are not metaphysically sufficient for perceptual experiences. A brain, stimulated in some specific way, is not, and never will be enough to create perceptual experiences. Here he is a disjunctivist of sorts, saying that when dreaming, for example, the experience one has relies on certain bodily processes (such as activity in the endocrine system), and produces a *dreaming* experience rather than a *perceptual* experience. Dream experiences, if primarily or solely brain-based, are not the same as the perceptual experiences that are essential to cognition, according to Noë. So any instance of brains in vats could entail a loss of perceptual experiences, and with it a loss of cognition for Noë (because cognition is an activity one does, not simply the activation of neurons). This would obviously mean that his theory would not allow for never-embodied brains to have cognition. However, is he opposed to the notion of a once embodied brain retaining cognition when disembodied?

He proudly states that his externalism is compatible with internalism's claim that "appropriate changes in the brain will produce appropriate changes in consciousness, even if the environment is unchanged." (Noë, 2004: 221) Meaning one could dream or hallucinate even when not interacting with the world. This may sound as if he could be falling into the same category as all others discussed so far, because the external environment is necessary for the providing perceptual experiences, yet conceptual experiences of other kinds can be retained by brain stimulation alone. But I don't think this puts him into the weaker camp.

The further point that reveals his support of the strong interpretation has to do with what allows an individual to experience anything at all. Aside from proprioceptive states one might experience in a dream that tie those thoughts to an embodied form, the emotive content relies on bodily processes from glands. Fear, joy, excitement, whatever—each of these emotions is not solely brain based. These provide experiential content and they are dependent on the physical substrate of the endocrine system. The brain needs input of some kind coupled with a body (of some kind) to generate any meaningful activity, be that direct electrical stimulation coupled with chemical processes from the glands, or fully embodied interaction with the world.[43] This is why he says neural activity alone is not sufficient for perceptual experiences and why a fully abstracted language of thought is not possible in his framework. The symbols may contain some information, but without some bodily engagement there will be no intentionality. One could allow that a neural-based language of thought has some content, but he would ask, "In what way is that content meaningful?" And the only way for that question to be answered is to consider the body of that agent, according to Noë. Additionally, he says a carefully stimulated brain showing some activity would prove little about the cognitive life of the brain. "At most it would show that experience could be produced by means of interaction between a probing scientist and a healthy animal." (Noë, 2004: 211) In other words, there is minimal worldly interaction with a minimally embodied agent, possibly resulting in an exceptionally weak form of "cognition" that would fall well short of the robust human cognition resulting from skilled sensorimotor activity, and would tell us little to nothing about the stimuli and experiences necessary for full human cognition.

Lakoff & Johnson

Metaphors are the foundation for all conceptualization, according to Lakoff and Johnson. In *Metaphors We Live By* (1980) they state the two main roles of metaphor in cognition are: (1) Metaphors provide the boundaries—expansive and restrictive—for one's cognitive capabilities, and (2) the metaphors we live by are based in large part on the particular type of bodily

---

[43] He is assuming intentional content is essential to mental representations and cognition, which is debatable, but not a point I will discuss.

experience of which human beings are capable. These need a bit of unpacking to see where they land regarding metaphysical dependency between cognition and mind-body-word interaction.

In saying metaphors provide the boundaries it's not initially clear whether this is a claim only about semantic content or if it is meant to have syntactic implications as well. Do the bodies we happen to have provide certain types of content for mental symbols or do our bodies actually affect the functions and rules governing them as well? To use one of their examples, does the fact that human beings generally have hands allow for one to represent concepts such as GRASP, or does the fact that humans have sensorimotor processes coupled with hands allow for mental content (such as GRASP and grasping metaphors) to be exercised altogether? The former is only a nomological claim, the latter metaphysical. They are suggesting the latter.

First off, they would say the question above is not a good one because it assumes mental symbols. Lakoff and Johnson believe cognition to be biological and neural rather than a matter of modal or amodal symbols governed by algorithmic processes. Instead, concepts like GRASP are composed of combinations of neural pathways (cross-domain mappings) due to, and reflective of, the peculiarities of our bodies. (Lakoff, 2003) As the body interacts with the world, grasping objects, the concept of GRASPING is formed via a process similar to Glenberg's idea of meshing, but rather than linking linguistic and perceptual symbols, different sensorimotor, linguistic, and perceptual domains of the brain are linked, creating unique conceptual content derived directly from experience. Lakoff and Johnson say *all* concepts are built up from these linked domains/experiences and thus all concepts have meaning only in so far as they are derived from body-based experiences and metaphors. Consequently, GRASP will have the meaning it does by linking the sensorimotor processes of using one's hand to hold an object with the parts of the brain responsible for processing the sight and sound of the word "grasp." Likewise, any concepts built up from this, such as UNDERSTANDING, will be due to the cross-domain mapping of GRASP as well as any other constituent metaphors and concepts. The mapping is similar to Hurley's horizontal modularity in how it literally constructs semantic content and inferential rules based on how the various domains are linked.

> The metaphor is not just a matter of language, but of thought and reason. The language is secondary. The mapping is primary, in that it sanctions the use of source domain language and inference patters for target domain concepts. The mapping is conventional, that is, it is a fixed part of our conceptual system, one of our conventional ways of conceptualizing [GRASP or UNDERSTANDING]." (Lakoff, 1993: 208)

When one conceptualizes something like GRASP, a linkage between these domains constitutes the meaning of GRASP and determines how the concept can be used in further cognitive tasks. Since GRASP is shaped and limited by the domains related to using one's hand, determining the size of external objects, the distance an object is from the body, etc., if one attempts to use "grasping" in a sentence applied to a concept or metaphor that shares no similar domains, the statement will be unintelligible. For example, the sentence, "Her voice grasped the heat" is confusing because HEAT and VOICE generally do not share the same domains as GRASP. Thus, there is no effective inference as to the meaning of the sentence because the metaphorical content is disassociated within the sentence. And, if it does make sense, it will be due to inferring body-based qualities. One might imagine the woman's voice with hand-like qualities, grasping warmed air in some way. This is also why anthropomorphizing is so prevalent in language—it allows us to conceptualize non-human entities in body-centric ways, grounding one's understanding of the world in uniquely human ways.

Like all embodied theorist discussed so far, if Lakoff and Johnson are correct, fully disembodied cognition—cognition in a mind that has never been embodied—is absurd because it removes any possibility of developing cognition. Embodiment is a metaphysically necessary constituent of cognition because there are no rules or representations that exist independent of embodied activity. This denial of mental symbols and innate syntax and the attempt to explain inference via cross-domain mapping is important because if they were only advancing a more elaborate argument for semantic content of mental representations, their argument and ideas could still be explained in terms of traditional cognitive science. A traditionalist could accept that bodily experience determines semantic content of mental symbols, and all content is derived in some way from this experience. This would be a nomological claim. Symbols have the meanings they have because of the bodies and sensory systems we happen to have. But a stronger claim is being made because the symbols and syntax are not physical constituents of the brain awaiting modal input that could be provided via embodied or disembodied stimuli. All of cognition is now dependent upon establishing neural maps shaped entirely by embodied interaction with the world.

However, their insistence on denying mental symbols may be a problem in terms of establishing continued embodied activity as a metaphysical necessity of cognition, because if cognition is based solely on cross-domain maps, it might be possible for maps to exist in disembodied brains. Since inference patterns emerge from and with neural conceptual maps, Lakoff and Johnson allow for the governing rules of cognition (whatever form those might take) to be entirely dependent upon the domains one possesses and the experiences or stimuli those domains receive. Thus, it should be possible for rules of cognitive patterns to develop in any cognitive system that is actively forming cross-domain maps. Lakoff and Johnson say embodiment is necessary because it forms the basis for all of our neural domains, which is why all of our concepts are derived from body-based metaphors. But this can be seen as a nomological issue. Our brains have those domains because they happen to be wired into a body with the sensorimotor systems humans happen to have. Because we have the hands we do, GRASPING is a foundational concept. Because we have only two sides, LEFT and RIGHT, and FRONT and BACK are foundational concepts as well. Their work provides groundbreaking insights into the source of humans' ability to conceptualize, yet all that has been successfully argued is that our body probably causes most of our conceptual content because these are the neural maps our embodiment produces. Lakoff and Johnson claim agents with very different bodies than a normal human will conceptualize the world in very different ways. The necessities for cognition are just that an agent must have cross-domain mapping from which concepts can develop and inferential rules/properties can emerge. But if cross-domain maps are the critical feature, they have not successfully shown some sort of mapping couldn't develop or remain in a disembodied brain.

Consider two scenarios: one is a brain grown in a lab or envatted just after birth so the agent/brain never experiences embodied life; another is a normal, bipedal human with fully functioning sensory and neural systems who, at the age to 20, suffers physical trauma so severe the brain must be envatted (for some reason this is the only salvageable organ). In the first, cognition will not be possible if Lakoff and Johnson are correct. Primarily because the brain needs to develop separate, specified neural domains based on the motor capacities of the agent. In an embodied agent this is obligatory because the brain must divide operations in order to manage the many appendages and/or sensory modalities of the body. If those domains do not separate from one another, cross-domain mapping is impossible and thus no conceptualization, and hence no cognitive activity, can occur. Now we're in the same territory as the other embodied theories regarding permanent disembodiment versus once-embodied-now-disembodied scenarios. But a traditionalist could challenge this and ask if a never-embodied brain could be wired up so as to receive stimuli akin to a computer simulation of the external world. The concepts may not have semantic properties we could understand, as embodied agents with the sensorimotor systems we happen to have, but there could still be concepts of some sort derived from the input received. However, I think Lakoff and Johnson could still say such a brain will not be able to cognize until the brain divides domains in order to create sensory modalities and processes of some kind. And we do not yet know if stimulation alone could be successful in achieving this the same way of embodied interaction. If so, then they may be willing to concede that a permanently disembodied brain, appropriately stimulated, could have cognitive activity. But until there is reason so suppose domains and cross-domain mapping can occur without embodied engagement with the world, they will likely deny it.

In the second scenario, based on Lakoff and Johnson's theory, the brain should still have cognition. The body has been lost, but so long as the brain is not significantly damaged, the neural maps should be intact. And if the neural maps are intact, the semantic content and inferential rules/patterns of cognition remain. Presumably this brain would still have very human-like cognitive activity because the foundational patterns were developed when the brain was in a human body and therefore the neural domains would have been defined according to human sensorimotor capabilities. And since the stimulation patterns of already established domains constitute cognition (as opposed to continuous input-output-input patterns, like Hurley argues) so long as the brain could be stimulated with similar mappings, or stimulations could be provided that would activate similarly mapped pathways, the same concepts would be utilized, the same metaphors articulated, and cognition itself would still function even though the brain would be envatted and disembodied. In this scenario, it might be appropriate to say cognition is *caused* by embodied activity, but it does not continue to exist or operate because of it. Yet again, this is the weak interpretation. Though the body causes cognition (and cognitive content), cognition is occurring without mind-body-world interaction.

However, just as with Hurley, one could argue that Lakoff and Johnson are making the strong claim by broadening and redefining what constitutes neural domains. Since they do not believe input is abstracted to any sort of arbitrary mental representations, perhaps the neural domains they describe should not be thought of in the same way as neural modules in traditional cognitive science. In traditional cognitive science, the domains are clearly mind/brain based and the body is the conduit for specific modes of input those modules will process and translate. From the outset, though mind and body are linked, the causal path of the mental sandwich separates the two; the symbols and syntax reside in the mind and the input/output travels from and through the body. In contrast, I believe Lakoff and Johnson's model links body and mind so that the two are not clearly separable.

Consider the example of the disembodied-from-birth brain above. Until the brain becomes intertwined with a body, there are no domains within the brain—it is pure cognitive possibility at that point. Only when it is intertwined with a body do the domains of the brain take shape. Only when one has hands does a domain develop capable of judging an object as graspable; only when one has eyes does a domain develop capable of seeing an object as nearer or farther than others; only when one has a body in the world does one develop a domain able to establish BIGNESS and SMALLNESS; etc. Cognitive capabilities are produced only when the brain receives patterns of stimulation from sensorimotor systems entangled with neural structures. And conceptualization only becomes possible when patterns repeat over time and engage multiple domains. But because these patterns are not translated, and the domains materialize due to embodiment, the domains, once established, could be considered part of the body and not just part of the mind. It might be possible to interpret Lakoff and Johnson as saying the domains basically become internal extensions of the external capabilities. The body could be said to appropriate the brain or vice versa. Without translating input into mental symbols governed by a syntax that operates or exists independent of the sensorimotor domains and stimulation patterns, there is no clear way in which to separate mind and body because bodily activity is mental activity, and mental activity is a unique pattern of bodily activity. Another way of saying this is sensorimotor activity doesn't cause mental activity, instead domain-specific neural stimulations are part of sensorimotor activity and mental activity is identical to those neural stimulations. Again, if translation into a language of thought were taking place, the division between stimuli and language would provide the boundary for mind and body. But Lakoff and Johnson deny any such translation.

Considering this, if we look back at the envatted-but-once-embodied brain from earlier, rather than considering it a disembodied brain, it could be considered a significantly truncated but still embodied mind. The domains, as internal extensions of the body, still provide a sort of embodiment for the agent. This is because the domains don't represent previously embodied activity; they are the end point of embodied activity. And if they can still be stimulated to result in the use of cross-domain concepts in cognition, that brain is still having (minimally) embodied activity. Basically, the brain, once embodied in an agent with sensorimotor capacities, cannot be disembodied. The only way to truly disembody the brain is to remove those domains, which would also end all possibility for cognition. Thus, cognition, in any form, in any possible world, requires neural domains that, once entangled, should be regarded as parts of the body.

This is a way we could interpret Lakoff and Johnson as endorsing the strong interpretation, but I'm not certain we should. They do not present their theory in this way—encouraging readers to blatantly blur the lines of brain and body once neural domains become mapped through embodied experience. But perhaps they should. In the final chapter, I'll discuss this a littler more as applied in general and not to Lakoff and Johnson alone.

So, what do we have now?

After all of this discussion it can finally be said that very few, if any, embodiment theorists are saying something wholly at odds with traditional cognitive science. Instead, there are a variety of theories with varied themes on the metaphysical relationship between cognitive activity and mind-body-world coupling. None—including traditional cognitive science—deny a metaphysical link between them. All readily admit that some semantic mental content metaphysically depends on an agent's embodied interaction with the external world. But after that, things get a little more complicated. Looking at the various theories, there are at least three distinct weak claims that some or all fit into.

1. If an agent doesn't interact with his/her environment, his/her brain wouldn't have developed properly and cognition might not be possible.
2. Without somatic concepts derived from embodied experience, an agent could not develop abstract concepts.
3. Without some (at least one-time) causal connection to the world, an agent could not have certain intentional states.

Every embodiment theorist discussed, and likely all traditional cognitive scientists, adopt the first claim. Some are much more detailed than others in describing how sensorimotor interaction shapes and parses the operations of the brain, but all admit the brain's functions and abilities are not pre-programmed. Bodily interaction with environment, at minimum, allows for unique brain development. So that is claim is not a surprise and is the most ubiquitous trait of theories of embodied cognition.

The second is clearly endorsed by Lakoff and Johnson, as they provide a thorough account as to how abstract concepts are derived from body-based metaphors grounded in an agent's embodied activity. Gibson would fit into this category too. Affordances are the basis for one's concepts and, since these exist out in the world, one must interact with the world in order to acquire the content necessary for somatic and abstract concepts. Glenberg would endorse this claim as well, for similar reasons, but he would also support the third.

Regarding the third, Glenberg's indexical hypothesis relies on perceptual symbols meshing with linguistic symbols to provide any sort of understanding. Thus entailing that an agent must have some causal connection with the world to generate intentional content. VTR would also support this claim, for slightly different reasons. They deny any one-to-one correspondence between percepts and natural objects, meaning the intentional content one derives from experience is unique to that agent and could not exist without some sort of embodied, causal interaction with the environment. Hurley, too, supports this position. Her model of horizontal modularity requires that an agent explore his/her sensorimotor contingencies in order to develop perceptual neural modules that will generate conceptual content. Lakoff and Johnson hold a similar view that cross-domain maps, caused by body-world interactions, constitute semantic content—so they fit into all three weak categories! But all of these stances are weak because they allow for possible situations wherein mental tokens can still function separate from the causes of their mental types, and cognitive capacities can still operate removed from any possible embodied interaction—a once embodied, now disembodied brain retaining cognitive abilities.

But only one philosopher discussed adopts the strong view and says that cognition can be exercised only if, there is continued mind-body-world interaction: Noë. He is fairly clear in his claims that a lack of bodily interaction with the world entails a loss of cognition because neural activity alone is not sufficient for the perceptual experiences that constitute cognition. Noë admits that a brain could be stimulated to create activity of some kind, but that activity will not be cognition in any meaningful sense. And here is where he adds a stronger criterion that separates him from others who agree with the third weak position: for Noë, intentionality cannot persist without mind-body interaction. I can't say for certain that others disagree with him on this point, but he has been the most overt at making this claim. Mental symbols of any kind only gain intentional content and intentional capabilities when tied to and receiving input from an agent's body. And meaningless data, input, or activity that would occur in a disembodied brain (once-embodied or permanently disembodied) isn't enough for cognition. By that I mean that the mark of the mental usually includes at least autonomy and awareness of other objects (and others) as distinct and separate from one's self. For Noë, these minimal traits cannot arise nor exist in an agent stripped of sensorimotor contingencies because these provide the meaning for our mental content. For example, a number has meaning insofar as it is a quantity of things, things part of, on, or near oneself (and oneself is based on one's shape and abilities). And if you were to abstract it away from quantity to just it's shape or the sound of what it's called, awareness of this is reliant on other sensorimotor capacities. And if somehow it were abstracted further, Noë would say we're again out of the realm of robust human cognition into a debate about neural activity alone.

And keep in mind that sensorimotor contingencies depend on an agent's ability to explore his/her environment in some way with the sensorimotor capacities. Many of the other disputes he has with traditional cognitive science (sensorimotor contingencies, cognitive tuning resulting from experience, lack of vertical modularity or linear causal pathways) on their own do not entail the strong interpretation of metaphysical dependence between cognitive activity and mind-body-world interaction. But when taken together, and with his arguments around what constitutes perceptual experience, and the essential role of the possibility of perceptual experience in overall cognition, there is no possibility in his framework, under any circumstances, where cognition is exercised in a disembodied brain.

But now what? If the dividing line for fundamentally separating embodied cognition and traditional cognitive science is defined by how a theory describes the metaphysical dependence between mind, body, and world, and Noë is the only one (mentioned here) who clearly supports the really radical claim contrary to traditional theories, what's the next step? The next and final section will address this in two ways. (1) Can Noë's claims be associated with another overarching theory or approach that would then also meet the strong criteria, thus giving us an indication of a broader group truly at odds with traditional theories? More specifically, I'll look at his theory as a version of relationalism, and if it is, whether that means all relationalist theories meet the strong criteria. And (2) I will discuss the issue brought up in this chapter with Noë and Lakoff and Johnson about whether it is nonsensical to discuss disembodied cognition in the sense that, if their theories are true, a mature embodied brain should be no longer be thought of as distinct from the body.

5. Is there hope for Noë's enactivism?

Noë's take on embodied cognition—specifically his version of enactivism—does seem to be making a claim fundamentally at odds with traditional cognitive science. He is claiming that *every* cognitive act occurs only if there is mind-body-world interaction. Thus, embodiment is necessary for the content of mental types as well as continued exercise of any resulting tokens or capacities derived from them. But now that Noë has been established as one of the few (maybe only) embodiment theorists making truly radical claims against the traditional model, a couple of other questions need to be addressed. (1) Is he, or his version of enactivism, alone? Or does his theory relate to any others (even ones that might live beyond cognitive science), such as relationalism? And (2) is his stance tenable?

In regards to the first question, finding others that align with Noë will not make or break his case against traditional cognitive science. But finding similar arguments and lines of thought might strengthen his case by providing more empirical and philosophical backing and possibly revealing new ways to recast his argument against the prevailing traditional models. And if his theory does align with or fit under the larger umbrella of relationalism, then perhaps all relationalist theories ought to be considered as rationale against traditional cognitive science.

In regards to the second question, the claim that cognitive capabilities cannot operate apart from the sensorimotor processes responsible for their content is radical even within the embodiment community. If Noë wants to effectively defend this claim, I believe he has to make the case that a mature embodied brain should be categorized as an extension of bodily sensorimotor systems. In this case, he could say that due to domain formations and mappings entangled with the body, there comes a point where there is no clear distinction between brain and body—to the point that even a "disembodied" brain should be considered a minimally embodied mind. If so, the once embodied now disembodied brain-in-a-vat scenario will not trouble Noë in the same way as it has others because the brain is never functionally disembodied. The term "functional" will be important. A way to understand Noë's view is as a sort of functionalism, where meeting the functional specifications requires a mind to be disposed to interact with the world through a body. Because a truly disembodied brain cannot manifest these dispositions it can't exercise (any) cognitive capacities. I don't believe he will be able to sustain his position against traditional cognitive science without establishing an entangled relationship between the brain and body. This was discussed briefly in the last chapter. Here I will delve deeper into the notion that an embodied brain, entangled with sensorimotor systems, should be thought of as an extension of the body rather than an independent organ.

Does Noë have an ally in relationalism?

Roughly speaking, relationalism is a branch of disjunctivism regarding perception that claims veridical perceptions of the world are constituted by mind-*independent* objects and properties. Non-veridical perceptions, such as hallucinations, are of different conscious kinds because they are constituted by mind-*dependent* properties (hence, disjunctivism). Where hallucinations may be solely brain-based—understood as the activation of mental representations and processes—veridical perceptions are made up of the relationship between a perceiver, an object, and the standpoint of the perceiver.

> According to the relational view of experience…we are not to think of perceptual consciousness of an object as a two-place relation between a person and an object, but rather as a three-place relation between a person, a standpoint, and an object…we need to factor in the notion of a 'standpoint' as our experience of

> objects is always in some sense partial. You always experience an object from a standpoint, and you can experience one and the same object from different standpoints. (Soteriou, 2014)

The idea of pulling in "standpoints" is in line with Noë's claims about sensorimotor contingencies and an agent's ability to traverse an environment and experience objects in several ways with the various senses. The standpoint of the agent, which can easily be seen as a result of an agent's position and sensorimotor capabilities, is a constituent of the perceptual experience because a change in position or perspective affects the qualitative nature of the perception. For now, these two approaches seem to agree with one another.

Additionally, this relationalist approach echoes some ideas promoted by dynamicists. Thelen and Smith stated, "The coherence [of the cognitive system] is generated solely in the *relationships* between the organic components and opportunities of the environment." (Thelen & Smith, 2006: 281; my emphasis) The opportunities to which Thelen and Smith were referring are affordances and sensorimotor contingencies, which could be thought of as an agent's "standpoint." For Noë, this is good news. This similarity provides a bridge to another line of theories in philosophy of mind that align with Noë's approach. And even more similarities are found when looking a little further into the reasons relationalism supports the notion of a three-place relation in veridical experiences.

Relationalists support the three-place relation between objects, agents, and standpoints as a solution to the following inconsistent triad:

(I) Physical objects are mind-independent.

(II) Physical objects are the direct objects of perceptions.

(III) The direct objects of perception are mind-dependent.

Most relationalists work to deny (III) because they believe it does not provide a satisfactory explanation as to how perceptions can provide semantic content that is about mind-independent objects rather than merely "concerning" them. (Robinson, 2012) But if we find the explanatory role in the relationship between agent, object, and standpoint, conscious experience can then connect us with external objects directly.

John Campbell, an advocate of relationalism, applies this reasoning to demonstrative terms like "this" and "that." He says the semantic value of these terms is justified and caused by our conscious experience. When we use "this" or "that" in any rationale involving objects, it necessarily involves reflecting on our conception of the object as mind-independent. These demonstrative terms assume a standpoint of the speaker or thinker existing separate from the object. If I discuss "that" chair, I am conceptually positing myself as distinct from the chair, regardless of whatever other properties the concept CHAIR has. If this notion of the mind-independence of objects is born out of mental representations of these objects, then those representations are presupposing the content of mind-independence without being grounded by anything more than other mental representations. And such an explanation is circular and not satisfactory. Conscious experience, he says, is the only way out of this circle.

> If one holds a view of experience according to which an experience is a mental state or event with an intentional content that represents the world as being a certain way…such a view of experience at best presupposes what is to be explained. It presupposes what is to be explained if the content of such a representational state is conceptual, and it doesn't have the resources to play the relevant explanatory role if its representational content is non-conceptual. What's required…is a conception of experience as something less than conceptual representation of the object of experience, but which nonetheless has the resources to explain how experience of an object can provide one with a form of knowledge of the semantic value of a demonstrative that refers to it. (Soteriou, 2014)

The only successful candidate, for Campbell, is the relational model.

This is very brief account of relationalism, but it's enough to reveal a key aspect that both relational and enactive models share. Unlike Chemero's theory of radical embodied cognition, or many dynamicists who want to rule out mental representations altogether, enactivists like Noë and relationalists both allow for mental representations to remain. But the representations never become abstracted to an amodal language of thought. Instead, the representations function specifically because they are tied to sensorimotor engagement with the world and relationships between perceivers and their percepts. And there may be some instances (such as veridical perceptual experiences) where mental representations play no role at all. For traditional cognitive science, if a process is cognitive it involves mental representations. For enactivism and relationalism, that's not always the case. A quick way to see how this works out, and when mental representations are involved and when/if they involve only neural processes was laid out by Pierre Steiner (2014). He provides a nice chart of how Noë's enactivist model views mental representations and how it compares to the traditional, computational model. (Steiner, 2014: 48)

Mental representations as… / ***Theory***	Made out of natural content	Necessarily symbolic, detailed, abstract	Necessarily intracranial	Necessarily involved in all cases of cognitive processing
Enactivism	Yes	No	No	No
Computational theory of mind	Yes	Yes	Yes	Yes

From this, it's easy to see where Noë's theory diverges from traditional cognitive science on several key points regarding mental representations. I believe if we were to do the same exercise with relationalism it would align exactly with enactivism. As a form of disjunctivism, it is making a strong claim about types of perceptual experiences being different than other types (veridical versus contemplative or hallucinatory experiences). Though veridical experiences do not rely on mental representations, contemplative or hallucinatory experience could, and those representations should be thought of as contentful physical structures. However, because veridical perceptual experiences do not require representations, and instead are due to the relationship between agent, object, and standpoint, relationalists deny the claims that mental representations are necessarily symbolic, abstract, intracranial, and involved in *all* cases of cognition. This alignment is impressive and thoroughgoing, and allows the following conclusion to be stated with confidence: Enactivists like Noë are relationalists. However, it would be wrong to state that all relationalists are enactivists. At least it can't be stated with confidence.

As discussed in the last chapter, Noë is making the case that every cognitive act involving symbols derived from embodied experience (which is likely all cognitive acts) occurs only if there is mind-body-world interaction. Relationalists are more ambiguous in their stance. They are saying that *some* cognitive experiences occur only if there is mind-body-world interaction. This definitely aligns with the weak interpretation (if mind-body-world interaction provides semantic content for mental symbols, then mind-body-world interaction is metaphysically necessary for cognition). And it could align with the strong interpretation Noë supports, depending on whether *some* cognitive experiences can eventually be expanded to *all* cognitive experiences. If the three-part relationship of object, agent, and standpoint is necessary for veridical perceptual experience, the relationalist model may be saying nothing more than the three-part relationship is metaphysically necessary for propositional states about the certainty of mind-independent objects. This is perfectly amenable to the weak interpretation of constitutivity because it is not making similar claims about all cognitive experiences or mental states. Thus, it cannot be said that relationalism entails enactivism entirely—though it does entail enactivism in regards to veridical experiences. It may be that as these theories are fleshed out more, the alignment will only become stronger and more apparent. However, for now, it is only safe to say enactivists are relationalists, but not necessarily the inverse.

Can Noë support the strong interpretation of constitutivity?

In the last chapter, when discussing Lakoff and Johnson, I introduced the idea of the brain being considered an extension of the body once domains and neural pathways have been mapped due to sensorimotor activity. If it is a possibility that an embodied brain should be considered part of the body, it will allow Noë to overcome the most obvious and significant challenge to his theory: the claim that mental content derived from sensorimotor interactions can function apart from the sensorimotor processes necessary for their instantiation. This point is difficult for Noë to overcome because there are obvious examples of cognitive acts such as dreaming or contemplation where sensorimotor activity is apparently not necessary for cognitive activity, and there are cases of sensorimotor content seeming to function even as the sensorimotor system is not present—such as phantom limbs after someone has lost an appendage and sensual memories when the sensorimotor system is damaged or no longer functioning.[44] Plus, it is just difficult for some to accept the notion that a once embodied, now disembodied brain could not continue to have cognitive function if it were preserved and stimulated in the right way. Noë is making this claim, but in order to do so he needs to explain how phantom limbs, and the like, work in enactivism—and that there is no possibility for a once embodied, now disembodied brain to cognize.

[44] Interestingly, I know a woman who has no sense of smell, due to a cyst blocking her olfactory bulb. This developed about 8 years ago. And though she cannot smell anything now, she still has strong memories of various tastes and smells—to the point that she sometimes thinks she tastes certain foods. However, she has no way of proving if the sensations are new or activations of taste and smell memories.

The key is in looking at how cognitive functions within the brain become organized due to sensorimotor activity. As mentioned before, until the brain becomes intertwined with a body the domains are mere dispositions. The occipital lobe will manifest the neural processes necessary for visual experience only if it is stimulated via the visual sensorimotor system in a certain way. It is pure cognitive possibility or potential at that point. Only when it is intertwined with a body do the domains of the brain definitively take shape. Cognitive capabilities are produced only when the brain receives patterns of stimulation from sensorimotor systems entangled with neural structures. And conceptualization only becomes possible when patterns repeat over time and engage multiple domains. Importantly, these patterns are not translated, according to Noë. The patterns are believed to be the direct result of sensorimotor activity and are not abstracted into amodal symbols with sensorimotor content. Once established, the patterns remains as a sort of electrochemical imprint on the brain, causing the various domains to materialize. Due to this kind of embodied imprinting, it might be appropriate to consider the domains as neural extensions of the body—as endpoints in the sensorimotor system—and not just parts of the brain. The domains basically become internal extensions of the external capabilities. The body can be said to appropriate the brain or vice versa as the two become intertwined in the development of sensorimotor systems. Another way of positioning this is that sensorimotor activity doesn't cause separate mental activity. Perhaps there is no point at which one can be clearly separated from the other and, instead, domain-specific neural stimulations become part of sensorimotor activity and mental activity is identical to those neural stimulations. Thus, one could say domain-designated mental activity is embodied activity. And since Noë, along with Lakoff and Johnson, is willing to say that mental content of all kinds is derived from sensorimotor activity, he is saying that every time a thought involves domain-specific activations (which is always), the thoughts are necessarily embodied because the domains are internal extensions of the body.

If the traditional model was correct and translations from the original sensorimotor processes into a language of thought were taking place, the translations would provide a boundary for mind and body. The point at which information is exchanged from one to the other would provide both a functional and physical dividing line between mind and body. The original sensorimotor processes could be considered extensions of the body, but any symbols translated further and stored in other, non-modal processes would be considered solely brain based. However, because Noë denies such a translation ever takes place that nice, clean separation between the function of the mind and function of the body is not present in his model.

If this is acceptable, the very idea of a once-embodied-now-disembodied brain no longer makes sense. If a brain has been embodied, and the neural domains defined, it still has sensorimotor systems and thus potential sensorimotor activity. The systems are significantly truncated but still present to some extent. The domains, as internal extensions of the body, still provide a sort of embodiment for the agent. This is because the domains don't just store symbols with information about previously embodied activity; they are the end point of an embodied sensorimotor system. And if they can still be stimulated to result in the use of cross-domain concepts in cognition, that brain is still having (minimally) embodied activity. Ultimately, the brain, once embodied in an agent with sensorimotor capacities, cannot be functionally disembodied. The only way to truly disembody the brain is to remove those domains, which would also end all possibility for cognition since that would remove all systems that imbue symbols with meaningful content. Thus, cognition, in any form, in any possible world, requires neural domains that, once entangled, can be regarded as parts of the body.

For possible evidence of this, one could look at the crucial period hypothesis—which says the ability to acquire language is linked to an ideal period in the biological age and cognitive development of an agent (Penfield & Roberts, 1959). The crucial period hypothesis says that if a person does not acquire language by a certain age, the domains of the brain and the cross-domain maps will be too firmly established to allow for the significant alteration that would come from, and are needed when introducing a natural language. From the standpoint of enactivism, this is saying that once neural pathways have been established (due to embodied activity in the world) the brain becomes so tightly mapped that it is as incapable of altering domains as an ear is incapable of becoming a nose.

Yet, even if this is true in regards to language, evidence of neural plasticity makes it a bit harder to posit the domains of the brain as becoming permanently assigned to specific sensorimotor processes. If a person who suffers damage to a part of the brain due to stroke, for example, can relearn motor and language skills by engaging other parts of the brain and developing new domain maps, it would seem that some mental tokens gained through sensorimotor experience can remain even when the processes necessary for their initial content are gone or have been significantly altered. And if so, this looks like good news for the traditional model. In stroke rehabilitation, it would appear that language acquisition, comprehension, and retention have been abstracted from the original neural processes and stored elsewhere in the brain. However, this is a point that can be challenged and rethought to show how it might be evidence that sensorimotor mental content is preserved in cases of neural plasticity specifically due to continued connections to sensorimotor systems.

Consider the Gradual Reorganization Principle presented by David Barrett (2013). Though he presents it for purposes other than the embodiment versus traditional cognitive science debate, it affords an opportunity for an embodiment theorist to show how plasticity might be evidence of the brain/mind as an extension of the body. The principle states, "the brain will reorganize neurological processes in response to degenerations of all kinds, thus saving psychological capacities, assuming sufficient time is allowed for the brain to adapt." (Barrett, 2013: 327) This principle is developed in response to a case study involving the serial lesion effect on albino rats from Patrissi and Stein (1975). In the study, the researchers found that some rats could retain memories of how to navigate a test maze even after significant portions of their frontal cortex had been removed. The rats who had large chunks of their brain removed at once were unable to remember how to navigate the maze, but those whose brains were altered in small, serial increments retained memories for much longer.

> Thus, the basic idea is that the brain can recover from fairly significant lesions to recover to a normal level of performance, assuming that the damage occurs over a suitable period of time. If the damage afflicting a particular area is sudden and occurs in a one-time fashion, the deficits are significantly worse than if the damage to the same area is gradual. More particularly, if the damage occurs gradually, the brain can find ways to implement the same psychological functions using whatever is left. (Barrett, 2013: 329-330)

This point is significant for Noë's case when considering the claim that semantic content derived from sensorimotor activity is retained only if the neural processes tied to sensorimotor interaction continue to function. Embodiment theories, and enactivism in particular, will accept that mental tokens with sensorimotor content can continue to function even when the input portion of that sensorimotor system, such as eyes and optic nerves for vision, are damaged. But if the neural processes are damaged, then the mental content will be damaged as well. On the other hand, traditional theories would say the mental symbols remain when sensorimotor systems are damaged because the content has been abstracted to an amodal language of thought and is not bound to sensorimotor neural processes. At first glance, plasticity might seem to indicate the brain's ability to reorganize mental content specifically because the content is in an amodal medium that does not rely on the areas of the brain necessary for the original content. Parts of the brain, along with the domain maps involving those parts of the brain are being removed, yet the memories or content created from that domain remains. As a simple illustration, it could be that portions of the frontal lobe were activated (and presumably) necessary when the rat first learned the route through the maze. Yet, even when significant portions of the frontal lobe were removed, the rat still retained the ability to traverse the maze. Thus, it would seem this cognitive ability and memory are not metaphysically dependent on continued access to the original neural sensorimotor processes.

However, a key component of neural reorganization for the rats—or stroke patients, or any other evidence of neural plasticity—is that the agents retain or regain the mental content only through additional sensorimotor activity. In the case of the rats, they were tested in the maze after each lesion. With each small lesion, there was likely damage to the mental content, yet each retesting provided a new opportunity for sensorimotor activity to find a slightly new neural pattern. The rats who lost a large portion of the brain all at once had little chance of retention because they lost most or all of the cross domain-mapping at once, rather than only a small portion of the map at a time—a portion small enough that it could be mostly or wholly replaced by physically going through the maze again. From an enactivist standpoint, translation of the sensorimotor content to an amodal language of thought is not what allowed the rats to remember how to traverse the maze; it was directly due to the continued sensorimotor interaction with the world that allowed cognitive abilities to remain even after a brain lesion. If the content were in an amodal language of thought, hypothetically the content could have been cut away with any lesion, or at different rates for each rat, because each lesion could have damaged the portion of the brain where it happened to reside.

Like every bit of scientific study presented over this entire debate, the serial lesion experiments are not definitive proof for or against embodiment. However, if plasticity can be shown to be the result of embodied activity—of the mind's ability to develop new maps due to continued embodied interaction with the world—then it could be further reason to believe the brain is molded by bodily engagement in such a way that the two should be considered as an entangled extension of a singular system.

I do believe there is good reason to pursue this line of inquiry and rationale, and not just as a matter of semantics in terms of what makes the body the body, the brain the brain, or the mind the mind. More research would need to be done focusing on how entangled the sensorimotor systems become with neural domains, and how metaphysically and nomologically tied the body is to the brain's function. I also believe this stance is vital for Noë or any other embodiment theorist. They need to definitively address why sensorimotor mental content cannot function independently from the sensorimotor processes (neural and/or extended) necessary for its instantiation. Without saying the mature, domain-specified brain is an internal extension of the body, the argument will always struggle to establish anything beyond the weak interpretation of saying mind-body-world interaction is metaphysically necessary for cognition because it provides the semantic content for mental symbols.

6. Conclusion

The goal of this work has been to disentangle the very cluttered field of embodied cognition. In pursuit of that goal, the primary themes in the field of embodied cognition have been examined, as well as why, and if, they offer significant challenges to traditional cognitive science models. This resulted in a sort of taxonomy of embodiment theories that allowed both the common threads and the disparate claims to become more evident. At the heart of most arguments in favor of embodied cognition are the following claims:

1. Mind, body, and world are "coupled"
2. Perception is perceptually-guided action
3. Mental representations (assuming they exist) are not amodal
4. Language is based on embodiment
5. Efforts in artificial intelligence support embodiment

The first is ever-present in embodiment theories, but the remaining four reveal the various branches and tactics embodiment supporters have utilized to challenge traditional cognitive science.

This taxonomy helped sift through the arguments and uncover a fundamental dividing line between embodiment theories: the metaphysical role of mind-body-world interaction in cognitive activity. All embodiment theorists believe there is no possible world in which cognition occurs without mind-body-world interaction. However, this is an ambiguous stance that could mean embodied interaction is necessary for mental content—which traditional cognitive science also supports—or that no cognitive activity can occur without mind-body-world interaction. Almost all embodiment theorists seem to be adopting the first, weaker interpretation. And the weaker interpretation fails to overcome the very difficult challenge of showing that mental content and capabilities derived from sensorimotor activity cannot continue to function independent from the sensorimotor processes necessary for their instantiation. This stance does not rule out the possibility of a brain retaining and exercising content derived from mind-body-world interaction even when the body is removed. Disembodied cognition, or a brain in a vat, is still possible in with the weak interpretation and thus still possible in the vast majority of embodiment theories. For most, the only caveat is that the brain must have been embodied at some point in order to cognize, because interaction between mind, body, and world is responsible for most or all mental content. However, once the content and capacities have been established, it might be possible for cognition to remain in a disembodied brain because the information and cognitive capabilities reside in the brain. Crucially, this is not at odds with traditional cognitive science. Fodor and others readily agree that the majority of our mental content could not exist without embodiment. They admit there would be nothing to think about without mind-body-world interaction. But having nothing to think about does not entail the inability to think at all.

Noë's version of enactivism is the only model discussed that overcomes this issue and argues for cognitive contents and capacities that can only exist and be exercised in an embodied agent. Since he sees perceptual experiences as essential to cognitive content and function, and these experiences are of a different, inimitable type than solely brain-based processes, cognitive activity ceases when mind-body-world interaction ceases. Hence, brains in vats cannot cognize. This is his stance, but one that is difficult to hold against the once embodied, now disembodied brain scenario. Even though perceptual experiences are inimitable and essential to cognition, it seems quite possible that this type of disembodied brain could still have cognitive activity of some kind. And this is a door his, or any embodied theory cannot leave open if the goal is to replace traditional cognitive science.

One solution may be in finding similarly disposed theories, such as relationalism, to bolster the claims of enactivism and firmly establish a cognitive model that demands continued mind-body-world interaction and reveals the non-essential role of an amodal language of thought. Another solution is solidifying the metaphysical role of mind-body-world interaction in cognitive activities by recasting the brain as an extension of the body. This is not something that can be done easily or lightly, but one I believe can succeed when building upon the already entangled relationship between bodily sensorimotor systems and neural sensorimotor processes.

Bibliography

Adams, F. and Aizawa, K. (2008)—*The Bounds of Cognition* (Malden, MA: Blackwell Publishing).

Adams, Z. and Maher, C. (2015)—"Cognitive Spread: Under What Conditions Does the Mind Extend Beyond the Body," *European Journal of Philosophy,* 23 (3), pp. 420-438.

Barrett, D. (2013)—"Multiple Realizability, Identity Theory, and the Gradual Reorganization Principle," *The British Journal for the Philosophy of Science*, 64, pp. 325-346.

Barsalou, L. (1999)—"Perceptual symbol systems," *Behavioral and Brain Sciences*, 22, pp. 577-660.

Barsalou, L. (2008)—"Grounded Cognition", *The Annual Review of Psychology*, 59, pp. 617-645.

Beer, R. (2003)—"The Dynamics of Active Categorical Perception in an Evolved Model Agent, *Adaptive Behavior*, 11, pp. 209-43.

Beer, R. (2014)—"Dynamical Systems and Embedded Cognition," in *The Cambridge Handbook of Artificial Intelligence*, Frankish, K. and Ramsey, W., eds. (Cambridge University Press), pp. 128-150.

Block, N. (2009)—"Comparing the Major Theories of Consciousness", in *The Cognitive Neurosciences IV*, M. Gazzaniga, ed. (Cambridge, MA: MIT Press), pp. 1111-1122.

Block, N. and Fodor, J. (1972)—"What Psychological States are Not", *Philosophical Review* (April), pp. 159-189.

Block, N. and Stalnaker, R. (1999)—"Conceptual Analysis, Dualism and the Explanatory Gap", *The Philosophical Review*, 108 (1), pp. 1-46.

Brooks, R. (1991)—"New Approaches to Robotics," *Science*, 253, pp. 1227-1232.

Brooks, R. and Stein, L.A. (1994)—"Building Brains for Bodies," *Autonomous Robots*, 1, pp. 7-25.

Brooks, R., Breazeal, C., Marjanovic, M., Scassellati, B., and Williamson, M. (1998)—"The Cog Project: Building a Humanoid Robot," in *Computation for Metaphors, Analogy and Agents*, Vol. 1562 of Springer Lecture Notes in Artificial Intelligence, C. Nehaniv, ed. (Springer-Verlag).

Caggiano, V., Fogassi, L., Rizzolatti, G., Casile, A., Giese, M., and Their, P. (2012)—"Mirror neurons encode the subjective value of an observed action," *PNAS*, 109, pp. 11848-11853.

Casile, A., Caggiano, V., and Ferarri, P. (2011)—"The Mirror Neuron System: A Fresh View," *The Neuroscientist*, 17, pp. 524-538.

Chalmers, D. and Jackson, F. (2001)—"Conceptual Analysis and Reductive Explanation", *The Philosophical Review*, 110 (3), pp. 315-360.

Chalmers, D. (2003)—"Consciousness and its Place in Nature", from *Blackwell Guide to the Philosophy of Mind*, S. Stich and F. Warfield, eds. (Malden, MA: Blackwell, 2003), pp. 102-142.

Chemero, A. (2009)—*Radical Embodied Cognitive Science* (Cambridge, MA: MIT Press).

Churchland, Patricia and Sejnowski, T. (1990): "Neural Representation and Neural Computation", *Philosophical Perspectives*, 4, pp. 343-382.

Churchland, Patricia (2011)—*Braintrust* (Princeton, NJ; Princeton University Press).

Churchland, Paul (1981)—"Eliminative Materialism and the Propositional Attitudes", *The Journal of Philosophy*, 78:2, pp. 67-90.

Clark, A. (1997)—*Being There: Putting Brain, Body, and World Together Again* (Cambridge, MA: MIT Press).

Clark, A. (2001)—"Visual Experience and Motor Action: Are the Bonds too Tight?", *The Philosophical Review*, 110, pp. 495-519.

Clark, A. (2011)—*Supersizing the Mind: Embodiment, Action, and Cognitive Extension* (New York: Oxford University Press).

Clark, A. and Chalmers, D. (1998)—"The Extended Mind", in *Mind and Cognition: An Anthology*, Clark, A. and Toribio, J. (2001)—"Sensorimotor Chauvinism: Commentary on J.K. O'Reagan and A. Noë 'A sensorimotor account of vision and visual consciousness'", *Behavioral and Brain Sciences*, 24:5.

Cowart, M. (2005)—"Embodied Cognition," Internet Encyclopedia of Philosophy, Fieser, J. and Dowden, B., eds., http://www.iep.utm.edu/embodcog/.

Cummins, R. (1989)—*Meaning and Mental Representation* (Cambridge, MA: MIT Press).

Dennett, D. (1992)—*Consciousness Explained* (Boston: Black Bay Books).

Domasio, A. (1989)—"Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition," *Cognition*, 33, pp. 25-62.

Feron, J., Gentaz, E. & Streri, A. (2006)—"Evidence of amodal representation of small numbers across visuo-tactile modalities in 5-month-old-infants," *Cognitive Development*, 21, pp. 81-92.

Fodor, J. (1974)—"Special Sciences", *Synthese*, 28, pp. 77-115.

Fodor, J. (1975)—*The Language of Thought* (Cambridge, MA: Harvard University Press).

Fodor, J. (1978)—"Propositional Attitudes", *The Monist*, 61, pp. 501-523.

Fodor, J. (1983)—*The Modularity of Mind* (Cambridge, MA: MIT Press).

Fodor, J. (2008)—*LOT 2: The Language of Thought Revisited* (Oxford: Oxford University Press).

Fodor, J. and Pylyshyn, Z. (1998)—"Connectionism and cognitive architecture: A critical analysis *Cognition,* 28, pp. 3-71.

Fogassi, L. and Ferrari, P. (2012)—"Cortical Motor Organization, Mirror Neurons, and Embodied Language: An Evolutionary Perspective," *Biolinguistics*, 6, pp. 308-337.

Gallagher, S. (2006)—*How the Body Shapes the Mind* (Cambridge: Oxford University Press).

Gallese, V., Fadiga, L., Fogassi, L. and Rizzolatti, G. (1996)—"Action recognition in the premotor cortex," *Brain*, 119, pp. 593-609.

Gallese, V. (2007)—"Before and below 'theory of mind': embodied simulation and the neural correlates of social cognition", *Philosophical Transactions: Biological Sciences*, 362, pp. 659-669.

Garbarini, F. and Adenzato, M. (2004)—At the root of embodied cognition: Cognitive science meets neurophysiology," *Brain and Cognition*, 56, pp. 100-106.

Glenberg, A. (1997)—"What memory is for," *Behavioral and Brain Sciences*, 20, pp. 1-55.

Glenberg, A. (2006)—Naturalizing Cognition: The Integration of Cognitive Science and Biology," *Current Biology*, 16, pp. R802-R804.

Glenberg, A. and Kaschak, M. (2002)—"Grounding Language in Action," *Psychonomic Bulletin & Review*, 9, pp. 558-565.

Glenberg, A. and Robertson, D. (2000)—"Symbol Grounding and Meaning: A Comparison of High-Dimensional and Embodied Theories of Meaning," *Journal of Memory and Language*, 43, pp. 379-401.

Glenberg, A., Havas, D., Becker, R., and Rinck, M. (2005)—"Grounding Language in Bodily States: The Case for Emotion," in R. Zwaan and D. Pecher, eds., *The Grounding of Cognition: The Role of Perception and Action in Memory, Language, and Thinking* (Cambridge: Cambridge University Press), pp. 115-128.

Goldin-Meadow, S. (2003)—*Hearing gesture: How our hands help us think* (Cambridge: Cambridge University Press).

Gonzalez, J. et al. (2006)—"Reading cinnamon activates olfactory brain regions," *Neuroimage*, 32, pp. 906-912.

Goodale, M. & Milner, D. (1992)—"Separate visual pathways for perception and action," *Trends in the Neurosciences*, 15 (1), pp. 20-25.

Grezes, J. et al. (2003)—"Objects automatically potentiate action: an fMRI study of implicit processing," *European Journal of Neuroscience*, 17, pp. 2735-2740.

Haugeland, J. (1981)—"Semantic Engines: An Introduction to Mind Design", in *Mind Design*, J. Haugeland, ed. (Cambridge: MIT Press), pp. 1-34.

Haugeland, J. (1993)—"Mind Embodied and Embedded", in *Mind and Cognition: 1993 International Symposium*, Y. Houng and J. Ho, eds. (Academia Sinica, 1993), pp. 207-237.

Hauser, M., Chomsky, N. and Fitch, W.T. (2002)—"The Faculty of Language: What Is It, Who Has It, and How Did It Evolve," *Science*, 298, pp. 1569-1579.

Hirosi, M. and Ogawa, K. (2007)—"Honda humanoid robots development," *Philosophical Transactions of the Royal Society A*, 365, pp. 11-19.

Horgan, T. and Kriegel, U. (2008)—"Phenomenal Intentionality Meets the Extended Mind", *The Monist*, 91, pp. 353-380.

Horgan, T. and Tiensen, J. (2002)—"The Intentionality of Phenomenology and the Phenomenology of Intentionality", in *Philosophy of Mind: Classical and Contemporary Readings*, D. Chalmers, ed. (Oxford: Oxford University Press, 2002).

Hotton, S. and Yoshimi, J. (2011)—"Extended Dynamical Systems Theory to Model Embodied Cognition," *Cognitive Science*, 35, pp. 444-479.

Hurley, S. (2001)—"Perception and Action: Alternate Views", *Synthese*, 129, pp. 3-40.

Hurley, S. (2008)—"The Shared Circuits Model: How Control, Mirroring, and Simulation can Enable Imitation, Deliberation, and Mindreading", *Behavioral and Brain Sciences*, 31, pp. 1-58.

Hurley, S. and Noë, A. (2003)—"Neural Plasticity and Consciousness", *Biology and Philosophy*, 18, pp. 131-168.

Kaschak, M. and Glenberg, A. (2000)—"Constructing Meaning: The Role of Affordances and Grammatical Constructions in Sentence Comprehension," *Journal of Memory and Language*, 43, pp. 508-529.

Keysers, C. and Gazzola, V. (2010)—"Social Neuroscience: Mirror Neurons Recorded in Humans," *Current Biology*, 20 (8), R353-R354.

Kiefer, M. (2005)—"Repetition priming modulates category-related effects on even-related potentials: further evidence for multiple cortical semantic systems," *Journal of Cognitive Neuroscience*, 17, pp. 199-211.

Kim, J. (1995)—"Mental Causation: What, Me Worry?", *Philosophical Issues*, 6, pp. 121-153.

Kim, J. (2006)—*Philosophy of Mind* (Cambridge, MA: Westview Press).

Lakoff, G. (1993)—"The contemporary theory of metaphor," in *Metaphor and Thought*, 2nd ed., Ortony, A., ed. (Cambridge: Cambridge University Press), pp. 202-251.

Lakoff, G. and Johnson, M. (1999)—*Philosophy and the Flesh: The Embodied Mind and its Challenge to Western Though* (New York: Basic Books).

Lee, H. & Vecera, S. (2005)—"Visual Cognition Influences Early Vision: The Role of Visual Short-Term Memory in Amodal Completion," *Psychological Science*, 16 (10), pp. 763-768.

Levinas, E. (1969)—*Totality and Infinity* (Pittsburgh: Duquesne University Press)

Lewis, D. (1972)—"Psychophysical and Theoretical Identifications," *Australasian Journal of Philosophy*, 50 (December), pp. 249-258.

Lewis, D. (1988)—"What Experience Teaches," in *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*, P. Ludlow, Y. Nagasawa, and D. Stoljar, eds. (Cambridge, MA: MIT Press, 2004), pp. 77-104.

Lyons, J. (2001)—"Carving the Mind at its (Not Necessarily Modular) Joints," *The British Journal for the Philosophy of Science*, 52, pp. 277-302.

Maravita, A., and Iriki, A. (2004)—"Tools for the body (schema)," *Trends in Cognitive Science*, 8 (2), pp. 79-86.

Marr, D. (1982)—*Vision* (San Francisco: Freeman).

Martin, A. (2007)—“The Representation of Object Concepts in the Brain,” *Annual Review of Psychology*, 58, pp. 25-45.

Merleau-Ponty, M. (1962)—*Phenomenology of Perception* (New York: Routledge).

Merleau-Ponty, M. (1963)—*The Structure of Behavior* (Beacon Press).

Miceli, G., Fouch, E., Capasso, R., Shelton, J., Tomaiuolo, F. & Caramazza, A. (2001)—“The dissociation of color from form and function knowledge,” *Natural Neuroscience*, 4 (6), pp. 662-667.

Millikan, R.G. (1986)—“Biosemantics”, The Journal of Philsophy, 86 (6), pp. 281-297.

Newell, A. and Simon, H.A. (1961)—“Computer Simulation of Human Thinking,” Science, 134, pp. 2011-2017.

Noë, A. (2004)—*Action in Perception* (Cambridge, MA: The MIT Press).

Noë, A. and Thompson, E. (2004)—“Are there neural correlates of consciousness?”, *Journal of Consciousness Studies*, 11, pp. 3-28.

Noë, A. (2012)—*Varieties of Presence* (Cambridge, MA: Harvard University Press).

Novich, L. and Cheng, P. (2004) “Assessing Interactive Causal Influence,” *Psychological Review*, 111 (2), pp. 455-485.

Ochs, E., Gonzales, P., and Jacoby, S. (1996)—“When I come down I’m in the domain state”: Grammar and graphic representation in the interpretive activity of physicists, from *Interaction and Grammar* (Cambridge University Press: New York), pp. 328-329.

O’Regan, J. and Noë, A. (2001)—“A Sensorimotor Account of Vision and Visual Consciousness,” *Behavioral and Brain Sciences*, 24, pp. 939-1031.

O’Regan, J., Myin, E., and Noë, A. (2004)—“Towards and Analytic Phenomenology: The Concepts of ‘Bodiliness’ and ‘Grabbiness’”, in *Seeing, Thinking and Knowing: Meaning and Self-Organization in Visual Cognition and Thought*, A. Carsetti, ed. (Norwell, MA: Kluwer Academic Publishers, 2004), pp. 103-114.

Penfield, W. & Roberts, L. (1959)—*Speech and Brain Mechanisms* (Princeton, NJ: Princeton University Press).

Pezzulo, G., Barsalou, L. et al (2013)—“Computational Grounded Cognition: a new alliance between grounded cognition and computational modeling,” *Frontiers in Psychology*, doi: 10.3389/fpsyg.2012.00612.

Poggio, T. (1981)—“Marr’s computational approach to vision,” *Trends in Neuroscience*, 4 (10), pp. 258-262.

Polger, T. (2002)—“Putnam’s Intuition”, *Philosophical Studies*, 109 (2), pp. 143-170.

Putnam, H. (1975)—“The Nature of Mental States”, in *Mind and Cognition: An Anthology*, W. Lycan and J. Prinz, eds. (Malden, MA: Blackwell, 2008), pp. 40-47.

Pylyshyn, Z. (2007) – *Things and Places: How the Mind Connects with the World* (Cambridge, MA: MIT Press).

Raibert, M., Blankespoor, K., Nelson, G., Playter, R. (2008)—“BigDog, The Rough-Terrain Quadruped,” http://www.bostondynamics.com/img/BigDog_IFAC_Apr-8-2008.pdf.

Rauscher, F., Krauss, R. and Chen, Y. (1996)—“Gesture, Speech, and Lexical Access: The Role of Lexical Movements in Speech Production,” *Psychological Science*, 7 (4), pp. 226-231.

Reboul, A. (2012)—“Radical Embodied Cognition vs. ‘Classical’ Embodied Neuroscience”, in *Cognition and Embodiment*, Z. Yu, ed. (Shanghai University Press).

Richardson, M.J. and Chemero, A. (2014)—“Complex Dynamical Systems and Embodiment,” in *The Routledge Handbook of Embodied Cognition*, Shapiro, L., ed. (New York: Routledge), pp. 39-50.

Robinson, H. (2012)—“Relationalism versus Representationalism: How Deep is the Divide?”, *The Philosophical Quarterly*, 62 (248), pp. 614-619.

Rowlands, M. (2010) – *The New Science of the Mind: From Extended Mind to Embodied Phenomenology* (Cambridge, MA: MIT Press).

Roy, J. (1999)—“Saving Intentional Phenomena: Intentionality, Representation, and Symbol”, in *Naturalizing Phenomenology: Issues in Contemporary Phenomenology and Cognitive Science*, J. Petitot, F. Varela, B. Pachoud, and J.M. Roy, eds. (Stanford, CA: Stanford University Press, 1999), pp. 111-147.

Rupert, R. (2012)—“Representation in Extended Cognitive Systems: Does the Scaffolding of Language Extend the Mind,” in *The Extended Mind,* Menary, R., ed. (Cambridge, MA: MIT Press), pp. 325-354.

Sambre, P. (2012)—“Fleshing out Language and Intersubjectivity: An Exploration of Merleau-Ponty’s Legacy to Cognitive Linguistics,” *Journal of Cognitive Semiotics*, 4, pp. 189-224.

Searle, J. (1980)—"Minds, Brains, and Programs." *Behavioral and Brain Sciences*, 3, pp. 417-424.

Searle, J. (1992)—*The Rediscovery of the Mind* (Cambridge, MA: MIT Press).

Seifert, L. (1997)—"Activating Permanent Representations in Memory: Different Benefits for Pictures and Words," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23 (5), pp. 1106-1121.

Shapiro, L. (2011)—*Embodied Cognition* (New York: Routledge).

Simmons. W. K., Martin, A. & Barsalou, L. (2005)—"Pictures of Appetizing Foods Activate Gustatory Cortices for Taste and Reward," *Cerebral Cortex*, 15, pp. 1602-1608.

Simon, H.A. and Newell, A. (1971)—"Human problem solving: The state of the theory in 1970," *American Psychologist*, 26 (2), pp. 145-159.

Soteriou, M. (2014) – "The Disjunctive Theory of Perception", *The Stanford Encyclopedia of Philosophy*, Zalta, E., ed., http://plato.stanford.edu/archives/sum2014/entries/perception-disjunctive/.

Sprevak, M. (2009)—"Extended cognition and functionalism," *The Journal of Philosophy*, 106, pp. 503-527.

Steiner, P. (2014)—"Enacting anti-representationalism: The scope and the limits of enactive critiques of representationalism," *Avant*, 5 (2), pp. 43-86.

Stewart, J., Gapenne, O., Di Paolo, E. A. (2011) – *Enaction: Towards a New Paradigm for Cognitive Science* (Cambridge, MA: MIT Press).

Symes, E., et al (2008)—"Grasp preparation improves change detection for congruent objects," Journal of Experimental Psychology: Human Perception and Performance, 34 (4), pp. 854-871.

Theios, J. & Amrhein, P. (1989)—"Theoretical analysis of the cognitive processing of lexical and pictorial stimuli: Reading, naming, and visual and conceptual comparisons," *Psychological Review*, 96, pp. 5-24.

Thelen, E. & Smith, L. B. (1994) – *A dynamic systems approach to the development of cognition and action* (Cambridge, MA: MIT Press).

Thelen, E. & Smith, L. B. (2006)—"Dynamic Systems Theories," in *Handbook of Child Psychology*, Volume 1, Theoretical Models of Human Development, 6th edition, Damon, W. and Lerner, R. eds. (Wiley), pp. 258-312.

Thompson-Schill, S. L. (2003)—"Neuroimaging studies of semantic memory: inferring 'how' from 'where'," *Neuropsychologia*, 41, pp. 280-292.

Tipper, S., Paul, M., & Hayes, A. (2006)—"Vision-for-action: The effects of object property discrimination and action state on affordance compatibility effects," Psychonomic Bulletin & Review, 13 (3), pp. 493-498.

Tucker, M. & Ellis, R. (2001) – "On the Relations Between Seen Objects and Components of Potential Actions," *Journal of Experimental Psychology: Human Perception and Performance*, 24 (3), pp. 830-846.

Tversky, B. (2009)—"Spatial Cognition: Embodied and Situated," in *Cambridge Handbook of Situated Cognition*, Robbins, P. and Aydede, M., eds. (Cambridge: Cambridge University Press), pp. 201-216.

Vainio, L., et al (2008)—"On the relations between action planning, object identification, and motor representations of observed actions and objects," Cognition, 108 (2), pp. 444-465.

Van Gelder, T. (1995)—"What Might Cognition Be, If Not Computation," *The Journal of Philosophy*, 92, pp. 345-381.

Varela, F., Thompson, E., and Rosch, E. (1991)—*The Embodied Mind* (Cambridge, MA: MIT Press).

Wheeler, M., Petersen, S., and Buckner, R. (2000)—"Memory's echo: Vivid remembering reactivates sensory-specific cortex," *PNAS*, 97 (20), pp. 11125-11129.

Wilson, A.D. and Golonka, S. (2013)—"Embodied cognition is not what you think it is," *Frontiers in Cognitive Psychology*, 4 (58). doi: 10.3389/fpsyg.2013.00058.

Wilson, R. A. and Foglia, L., (2011) – "Embodied Cognition", *The Stanford Encyclopedia of Philosophy*, Zalta, E., ed., http://plato.stanford.edu/archives/fall2011/entries/embodied-cognition.

Zahavi, D. (2002)—"First-person thoughts and embodied self-awareness. Some reflections on the relation between recent analytical philosophy and phenomenology," *Phenomenology and the Cognitive Sciences,* 1, pp. 7-26.