2007

Somatic computationalism: Damasio's clever error

Christopher D. Pope
Louisiana State University and Agricultural and Mechanical College

SOMATIC COMPUTATIONALISM:
DAMASIO'S CLEVER ERROR

A Thesis

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Master of Arts

in

The Department of Philosophy and Religious Studies

by
Christopher D. Pope
B.A., Mississippi State University, 2000
M.A., Mississippi State University, 2002
August 2007

TABLE OF CONTENTS

ABSTRACT

Neuroscientist Antonio Damasio wrote a book entitled *Descartes' Error* (1994) in order to address popular misconceptions about the mind, particularly those which relate to Cartesian philosophy. One of the author's major goals for the book is to argue that emotion contributes to reason, that emotion is in fact necessary for rational thought to occur. In order to link emotion to reason, Damasio proposes a theory of mind which explains several mental functions in terms of neurological representations. Consciousness, reason, instinct and emotion all occur because the brain forms representations of the subject's body and of the world in which the body acts. Thought, in the broadest sense of the term, is the process in which the brain manipulates these representations and causes them to interact.

This thesis will examine Damasio's theory of mind in relation to two traditional topics in cognitive science: consciousness and intelligence. The first chapter simply explains the theory as given in *Descartes' Error*. Chapter two argues that, like everyone before him, Damasio fails to explain how or why the brain generates consciousness. Although the theory fails in this regard, it is still useful as a description of the neurological processes which underlie consciousness, of the mechanics of mind. As such, this theory could serve as a conceptual complement to the traditional paradigms of cognitive science, “GOFAI” and Embodied Cognition. Chapter three will argue that Damasio's theory is better suited to work with the latter paradigm than the former.

CHAPTER ONE: DAMASIO'S ERRORS AND *DESCARTES' ERROR*

1.1 Introduction: Damasio's Errors

After a long clinical and research career, neurologist Antonio Damasio wrote *Descartes' Error* (1994) to correct certain popular misconceptions about the mind. As the title suggests, the author's aim was to undermine what he understood to be Cartesian ideas, ones which have influenced both popular and scientific beliefs about the mind. Specifically, Damasio first addresses mind-body dualism by discussing neurological evidence that brain states affect mental states. Secondly, he attempts to undermine what he calls the "high reason" tradition of philosophy, in which emotion and reason are treated as separate phenomena, with reason the superior faculty. In this second venture, Damasio winds up arguing that emotion is a necessary component of ostensibly rational (logical) thought. He starts with the case of Phineas Gage in order to support both of these theses, but also draws on other historic and contemporary evidence.

Damasio acknowledges near the end of *Descartes' Error* that dualism and the high reason tradition trace back at least to Plato, and he further admits that he targeted Descartes just because the Frenchman promulgated certain ideas so successfully in the modern era. It is not Damasio's aim teach the history of philosophy, but to teach some science. In doing so, he proposes a computational-representational theory of mind with the primary thesis that mind is the result of the interactions in the brain between representations of the body and representations of objects external to it. He contends that these interactions give rise to logical thought, emotion, instinct, and even consciousness itself. Damasio's very ambitious theory of mind is the topic of this treatise.

Chapter one of is simply an exposition of the evidence, arguments and the theory of mind presented in *Descartes' Error*. Chapter two will examine Damasio's assertion that consciousness results from the interaction of body-representations and object-representations. As clever as the theory of mind is in other respects, it simply fails as a theory of consciousness, which is the central contention of chapter two. Damasio returns to his theory of consciousness in *The Feeling of What Happens* (1999), but does not add any ideas which adequately explain consciousness, as shall also be discussed in the chapter two.

Chapter three will examine Damasio's theory from a computational perspective. Two of the major paradigms in cognitive science are "GOFAI" (for "Good Old Fashioned Artificial Intelligence") and "Embodied Cognition." The former paradigm asserts that the mind is the result of strictly logical operations acting on representations of objects and representations of the world at large. The latter paradigm, embodied cognition, is an alternative approach developed by robotics experts frustrated with the many failures of GOFAI. Cogburn and Megill (2005) have suggested that Damasio's theory would be a useful conceptual complement to GOFAI. They argue that adding emotion to computation might solve the infamous "frame problem" of cognitive science. However, chapter three of this thesis contends that Damasio's theory of mind would be a better complement to embodied cognition. Damasio's theory holds that the body is the central topic of mental representation. It seems more likely that somatic representation (and computations based on such) would solve the problems faced by embodied cognition than those faced by traditional GOFAI.

Damasio misunderstood certain philosophical concepts in his work, thence "Damasio's Errors." The first such error is that he only understands Cartesianism to include mind-body dualism and the "high reason" tradition. Because he does not understand Descartes in greater detail and is not familiar with the history of philosophy, he makes two more mistakes which are rather Cartesian in character. As discussed in chapter two, Damasio posits an idea rather like Descartes' homuncular theory of consciousness. Through a mistake in rhetoric, Damasio unintentionally presupposes consciousness on the part of unconscious brain representations, all in an attempt to explain consciousness. He unwittingly begs the question of the origin of consciousness. Also, as discussed in chapter three, Damasio posits a theory of mind which is computational (though it is based on somatic representations rather than logical operations, thus is not "computational" in the strictest sense of the term). Computationalism, the belief that human thought is composed of logical operations, is an outgrowth of the very "high reason" tradition which Damasio tries to undermine. It is thus somewhat strange for Damasio to propose a (somatic-) computational theory of mind as an alternative to "reason."

On the one hand, these three errors spring from a lack of philosophical sophistication on Damasio's part. On the other, one of those errors turns out to be rather inspired. The somatic computationalism which he suggests, but does not elaborate upon in detail, might lead to a new way of thinking about cognition. Specifically, somatic computation acting in tandem with embodied cognition might provide a new paradigm for cognitive science. Thus, while Damasio's work is based on some conceptual oversights, his errors are fortunately very instructive ones.

1.2 *Descartes' Error*

Damasio presents *Descartes' Error* as a refutation of the Cartesian philosophy of mind. However, Damasio does not explicitly identify which Cartesian concepts are his targets until the last chapter of the book. When Damasio does finally explain why he attacks Descartes, he admits that he might as easily have challenged Plato. He chose Descartes as the representative of a "high reason" tradition which includes Plato, Kant, and probably most philosophers in between. The author only aims at Descartes because the Frenchman promulgated dualism so successfully, because he is the source of much modern belief about the mind. Damasio uses "Cartesian" as an umbrella term which refers to any sort of dualistic thought, whether it comes from Descartes or merely resembles his work (Damasio 250).

Damasio attributes four general theses to Descartes. (1) There is an immaterial mind which exists separately from the body. (2) This mind (and not the brain) is the source of reason, including social and ethical cognition. (3) Reason and emotion are separate faculties of the mind; reason is superior and should thus dominate emotion. (4) In addition to mind-body dualism, there is something of a brain-body dualism as well. The brain and body proper are almost autonomous of each other. In this line of thought, the body merely provides life support for the brain, which might as easily live in a vat by itself. Damasio takes aim at all four of these ideas, albeit he never untangles them explicitly. Damasio challenges not only Descartes on these four theses, but also the wealth of scientific thought and common sense supported by dualism.

Damasio presents three interconnected theses in his challenge to Cartesianism. (A) Emotion, feelings and regulatory biological processes all participate in reason. (B)

Feelings are not ineffable qualities of an immaterial mind, but rather the brain's perception of the state of the body. The mind is the function of the brain in this thesis. (C) The body is the chief subject of mental representations. The "body proper," which is the body excluding the brain, is the frame of reference for mental operations. Taken together, these theses deconstruct the portion of Descartes' philosophy of mind detailed in (1) through (4) above.

Damasio never explicitly spells out his position on the relationship of mind to brain, but his attitude conforms to identity theory. He does not grant any quarter to the idea that mental states could be anything besides brain states. The elusive mental quality of brain processes simply is not a topic of discussion in his work. He never directly attacks dualism by, say, raising the problem of mind-brain interaction. However, it is implicit to his work that mental states just are brain states. He is so eager to prove that changes to the brain cause changes to the mind that he neglects to explicitly state a position on the mind-brain relationship. If he is not exactly an identity theorist, the differences between his attitude and identity theory are unnoticeably trivial.

His three theses lead to what he terms the "somatic marker hypothesis," which is a proposal for integrating emotion into an explanation of rationality. In this thesis, because emotions are indications of body state, emotions are able to "mark" various stimuli according to how they affect the body. Stimuli which affect the body in a pleasant way are marked in so that they become priorities in mental calculations, and negative stimuli are marked so that they can be avoided.

1.3 Evidence against Descartes

Damasio opens the book with four chapters describing the neurological evidence from which he will draw in later argumentation. He actually works from a somewhat small body of evidence, consisting of a handful of historic cases, anecdotal evidence of larger groups of brain-damaged people, a dozen cases from his own research, and evidence from animals. Damasio chose to present this evidence in narrative form, telling the stories of various brain-damaged people like scenes in a novel. While his empirical methods are not currently under discussion, it is worth mentioning that he bases the book on a somewhat small body of information. Nonetheless, his results seem reliable, and there is no mention in literature of empirical problems with his work.

Damasio starts with a chapter about the legendary Phineas Gage, somewhat surprisingly. This almost mythic case is usually cited as early evidence that brain damage changes a person's mind. It is somewhat surprising that Damasio is able to draw on this specific case for his thesis. Damasio finds in Gage's story evidence for his own central thesis, that emotions contribute to the processes of reasoning. Gage's job was to use explosive powder to clear rocks out of the paths of new railroad tracks. Someone would first drill a hole in the ground, then Gage would partially fill the hole with gunpowder, insert a fuse, then top the powder with sand. The sand had to be packed into place to make sure the blast went down into the rock, and not back up into the air. Gage was purportedly a master of this delicate task, praised and valued by his employers for his mastery of the work (Damasio 4). He even had an iron tamping-rod custom-made to his specifications. One day in 1848, Gage was working, became distracted, and tamped the

explosive directly. This, of course, made the powder explode, which projected the tamping-rod up into Gage's skull.

Amazingly, Gage survived the wound and even the severe infection that followed. The account of Gage after the accident is somewhat sketchy because no one understood the neuro-psychology of the case at the time, because there simply was no way to document the wound internally, because Gage left the U.S. for a few years, and because the Civil War coincided with the last few years of his life. But the available evidence shows that while most of Gage's intellectual capacity and motor skills were unharmed, his personality changed drastically. More specifically, his social behavior and decision-making ability changed after the accident. He lost sight in the left eye, but not the right. He could still walk normally and use his hands dexterously, and he had no noticeable difficulties with speaking or the general use of language (Damasio 8). Somehow, the hole in his head did not destroy most of his "rational" capacities, nor his motor skills.

While Gage was spared obvious damage, something more subtle did occur. He became a much less pleasant person. Where he had purportedly been a man of "temperate habits" and much energy before the accident, he became something else afterward (ibid.). He began using extraordinary profanity in public settings, with no apparent regard for the comfort of others. He also became capricious in his character, formulating great plans and quickly abandoning them. He also became peculiarly boastful of his wound, eagerly displaying it and the tamping-iron which he kept with him for the rest of his life (a possible case of "collecting" behavior in contemporary terms). One doctor commented that the balance between his intellectual faculty and animal propensities" had been destroyed (ibid). His friends and acquaintances noticed this,

famously commenting that "Gage was no longer Gage." He also could not hold down work after the accident. Though he had been a self-sufficient employee before the accident, the railroad company dismissed him, and afterwards he drifted from job to job. He eventually died in the care of a relative. What Damasio reveals in his telling of the story is not just that Gage's personality changed because of the wound, but that certain functions of his mind changed as well. While he could still walk, talk and do physical labor like an ordinary man, some subtle functions of his mind were damaged. Specifically, his social behavior changed, as did his emotional restraint and his capacity to think in the long-term. These cognitive deficits are part of what Damasio later terms the "Gage matrix," and are the signature of brain damage such as Gage's.

Luckily enough, Gage's skull was preserved in a medical museum. In the second chapter of his book, Damasio presents speculation about the precise nature of the damage the iron rod inflicted. Some of his colleagues did very clever investigation using computer imagery (Damasio 23, 31). Their studies concluded that Gage likely did not suffer damage to the language and motor-control areas of the brain, which is consistent with the story in the first chapter. They also concluded that Gage's damage was local to the orbital (or ventromedial) prefrontal region. Ventromedial means "middle of the belly," prefrontal means "before the frontal." Thus, the damage was to the underside of the frontal lobes, somewhat behind the very front of the brain. This is the part of the brain roughly above and behind the eye sockets. Having connected the behavior of Gage to a specific type of wound, Damasio goes on to discuss a patient with damage to the same part of the brain and similar changes in behavior.

Chapter three focuses on the case of one of the author's own patients named "Elliot," presented by Damasio as "A Modern Phineas Gage" (the title of the chapter). With firsthand knowledge of Elliot's case, the author is able to expand his account of the role of the orbital prefrontal region in rationality, social cognition, and planning. The author does not disclose Elliot's specific job occupation, but explains that the patient had been employed in a "business firm" (Damasio 35). In that capacity, he had to perform commonplace clerical tasks such as organizing documents. In his thirties, Elliot developed a type of brain tumor that, while itself benign, was also very dangerous because it compressed the frontal lobes. When Elliot's tumor was removed, the damaged frontal-lobe tissue was extracted with it. Thus, the patient suffered damage possibly similar to what Gage experienced.

Before the tumor, Elliot had been a responsible employee, a family man, and a generally well-adjusted member of society. But his personality changed after the tumor's removal. He had to be prompted to go to work in the morning, he could no longer prioritize tasks, thus he seemed to suffer something like the "frame problem" so famous in cognitive science. He would focus on trivial tasks, or on trivial aspects of tasks, to the exclusion of his larger goals. He would also lose focus and, for example, decide to spend an entire day reading one of the documents he was supposed to be sorting. He also could not easily shift from one task to another. In addition to his work problems, Elliot became involved in a questionable financial scheme that led to bankruptcy, this despite warnings from friends. Even forewarned, he would not act rationally to avoid a disastrous investment. He also began exhibiting collecting behavior. Damasio belabors the Gage comparison quite enough in the book, but it does seem significant that Elliot could no

longer work and exhibited collecting. There are, of course, differences between this case and Gage's. Elliot did not use profanity and was not as generally "intense" as the railway worker (Damasio 36-38). The neuroscientist admits that he has no explanation for the differences between their cases. Minor variations in the location of the wound or variations in their backgrounds or general personality might account for the differences in their outcomes. While the differences are a bit worrisome, the similarities are conspicuous and support the author's contentions.

What is significant in this case is that Elliot retained a great deal of his cognitive capacity. He seemed competent enough to return to work, in fact he was initially denied disability benefits. He could speak and move about normally, and his intelligence seemed generally intact. What had changed was just his personality, in common parlance. Damasio put him through an exhaustive battery of intelligence and personality tests, and Elliot did quite well on them. According to the Weschler test, he had superior intelligence. He also did well on tests for attention, working memory, short-term memory, memories of past events, arithmetic, language, new learning, and perception. In short, what would normally be considered his "reason" was intact. He also tested normally on personality tests, such as the MMPI. All of this seemed incompatible with the fact that he could not function in the real world (Damasio 41).

Quite significantly, the author also put Elliot through a number of tests of reasoning and ethical judgement, and he did well. On these tests he was able to reason about ethical and financial matters correctly. On one test, he was able to generate options as possible responses to ethical and financial situations (Damasio 46). On another test, Elliot showed he could anticipate consequences generally. On yet another type of test, he

was able to invent effective ways of accomplishing social tasks. On a fourth type of test, Elliot showed he was aware of specifically social, interpersonal consequences. And on a fifth type of test, he was able to make moral judgements. This is all significant because the patient was able to perform in theory what he could not do in the real world. His store of background knowledge and theoretical reasoning capacities were intact, they just did not work in real-world situations.

Damasio winds down the story of this patient with some speculation about his case, hinting at the eventual explanation. Elliot had access to his background knowledge, was able to reason about complex matters, and did not suffer basic defects of intellect. Whatever the damage was, it seemed to "set in at the late stages of reasoning, close to or at the point at which choice making or response selection must occur." The answer would ultimately lie in another consequence of the tumor. In their interviews, the author noted that Elliot's emotions were very flat, that he suffered "shallow affect." He could describe important things without emoting, he seemed extraordinarily contained. Damasio does not go on about this in great detail, but speculates at the end of chapter three that somehow this emotional flatness eliminated Elliot's reasoning capacity. Thus, this contemporary case affirms what the Gage story suggests, that the orbital prefrontal region is involved in emotion and reasoning, and that emotions participate in reasoning in some way (Damasio 50-51).

After describing Elliot's case, Damasio goes on in chapter four to briefly survey other cases of prefrontal damage. He sketches four historical cases similar to those of Gage and Elliot. The first is a stockbroker, first studied in 1932, who had a brain tumor similar to Elliot's. His personal life and career also unraveled, and his modesty and

effectiveness in life were replaced by braggadocio and an inability to enact any of the elaborate plans he devised. He also suffered from shallow emotional affect, unable to feel anything even when considering his own tragic story.

A second case from the 1940s was studied by Wilder Penfield and colleagues, and focused on a sixteen year old child who suffered severe damage to his frontal lobes. The child's personality failed to develop normally afterward, and his social behavior deteriorated. The third case, also beginning in the 1940s, was an infant who suffered frontal lobe damage near birth. The child was not stupid, but his behavior was always abnormal and maladaptive. He was never able to hold down a job or otherwise function normally in society, including romantically. In both of these cases, patients were not flexible in their thinking; they were rigid and unimaginative in their daily activities. They were unable to plan, organize their effort, or work gainfully. They also boasted about themselves, could not interact socially and were not very interested in life, as though they also suffered shallow affect.

The fourth case Damasio mentions is not a specific person, but comes from the studies of patients who underwent the prefrontal leucotomy procedure starting in the 1930s. This procedure was the precursor to the infamous frontal lobotomies that were so massively abused in the next two decades. In the leucotomy, certain parts of the prefrontal brain areas are deliberately damaged in order to alleviate the severe anxiety that certain mental disorders cause. While the patients did experience less anxiety, and their intellectual faculties were intact afterwards, the procedure led to shallow affect in these people. They also manifested less creativity and decisiveness afterward. Like Elliot, they were unable to feel very much of anything afterwards (Damasio 54-61).

Damasio also includes in chapter four a brief discussion on evidence from animals, specifically monkeys. Despite million of years of evolutionary distance between humans and the other primate species, damage to the prefrontal areas of monkeys has a very similar affect to what happens in people. The wounded monkeys display less emotion on their faces, and presumably experience less emotion. Prefrontal damage also disrupts their social behavior, particularly grooming. They also could no longer reason their way through standard lab tests, which is possibly analogous to the breakdown of planning and reasoning in human patients.

Damasio also introduces a second type of brain damage and associated mental disorder in chapter four. The new topic he presents turns out to be an important secondary source of information. A disorder called anosognosia results from damage to the right (and not left) sensorimotor areas of the brain. Patients with this damage are partially paralyzed on the left sides of their bodies, unable to move their left arms, for example. Interestingly, these people do not know that they are paralyzed. They act and talk as though they were not injured. When confronted, they seem surprised that they cannot move their limbs. Some insist that they might have once been paralyzed, but no longer are. One patient had to look around to find her own limp arm, because she had so little sensation from it. That is the main problem of anosognosia: the victims cannot sense their physical defects automatically or quickly, and this never disappears in severe cases.

The way Damasio explains it, the human brains contain internal representations or maps of their bodies. In anosognosia, these maps are no longer receiving updated information from the body. Therefore, these patients are not aware of their own bodies.

While these cases are markedly different from prefrontal cases, there are also some similarities. Anosognosics are notoriously unworried about their condition, they do not seem too perturbed by their conditions. They suffer shallow affect like prefrontal patients. Anosognosics are also unable to theorize about the future or implement plans effectively, also like prefrontal patients (Damasio 62-70).

This discussion of anosognosia might feel somewhat random, especially because it is in the middle of chapter four's survey of the history of prefrontal patients. However, it does still deal with the same concepts. Because emotion is the perception of body states, any interruption to the perception of body states will affect emotion. The disruption of emotion seems to be the primary cognitive defect of the pre-frontal cases. Thus, anosognosia is important as a secondary source of information to support the theses Damasio posits based on the prefrontal cases.

All of this evidence sets the stage for a refutation of Descartes, at least as far as Damasio understands him. The immaterial mind is quite obviously dependent upon and determined by the material brain (contrary to thesis 1 above). Reason, traditionally seen as the purview of the mind, is a process that occurs in the brain. Reason is disrupted when the brain is damaged (contrary to thesis 2 above). Reason apparently depends on emotion for its success. Patients who could no longer experience emotion properly could also not reason properly (contrary to 3 above). Damasio does not yet explicitly attack Descartes in these first four chapters, but he sets the stage for the overturn of Cartesianism. Having thus presented his evidence, the author proceeds in the next chapters to explain what happens in the minds of prefrontal patients and anosognosics. In

doing so, Damasio hopes to offer a theory to replace the commonsense notions of mind that more-or-less trace back to Descartes.

1.4 Damasio's Argument

In chapter five, Damasio begins with "Assembling an Explanation" for the strange behavior of the patients he discusses in previous chapters. He offers a computational theory of mind. Faced with a social or personal problem, the mind must use both "broad-based knowledge and reasoning strategies to operate over such knowledge" (Damasio 83). The author also observes that emotions arise from the same sorts of brain processes that serve to regulate body functions. He also notes that the brain and body interact in a very complex manner (contrary to thesis 4 above), using both nerves and bloodstream chemicals to pass information to each other. The brain and body are a single, unified organism, not separate entities as dualism suggests. There is a coherent theory of mind in all of these facts, and it a computational-representational theory.

The first feature of his theory of mind is the three-part nervous system consisting of sensory nerves, brain proper, and motor nerves. In his formulation, brains are basically the bundle of nerves between sensation and action. Some nerves are attached to sensory apparatuses and carry information to the brain. Others nerves are attached to muscles and take their cues from the brain, causing motion of one sort or another. Brains are the nervous tissue, the neurons, between sensory and motor neurons. As evolution produced more and more complex brains, more neurons were "interpolated" or inserted between sense and motor neurons (Damasio 89). It is in these intermediate neurons that the operations of mind take place. But, observes Damasio, inserting a few neurons between sense organs and muscles does not automatically produce a mind. This is to say

that the mere presence of a brain is not sufficient to produce a mind. A mind is "the ability to display images internally and to order those images in a process called thought" (ibid.). This is the heart of his theory: the mind consists of images generated and manipulated by the brain. Thus, his theory might be called "image computationalism." The brain is the nervous tissue between sensation and action, and it is the imagery in the brain that generates mind. What happens is that the brain receives sensory input, then performs some sort of operations on sensory images and memories (which are stored images). The brain then selects a possible motor response from a "menu" of existing responses, or it can generate a new motor response. Damasio is quick to point out that that images exist in every sensory modality. There are thus visual, tactile, olfactory, and auditory images in the mind (Damasio 93).

Having introduced this theory of mind, Damasio makes his first explicit critique of Descartes. He explains that there is no single area in the brain in which all of the images from the different sensory modalities combine into a composite. There are areas where multi-modal images form, but these are nothing like first-person phenomenology. This is all to say that the evidence of neuroscience shows there is no "Cartesian theater" in the brain, thus the homuncular version of mind is a "false intuition" (Damasio 94-96, 100). The appearance of mental unity is more a matter of simultaneous occurrence of different brain functions. The unified mind is thus perhaps an *illusion* generated by the timing of multiple brain functions. Although each sensory modality has its own attention and working memory, global attention and memory appears concentrated in the prefrontal areas. This is to say that the manipulation of images that constitutes consciousness occurs in the prefrontal areas damaged in Gage and the other cases (ibid.).

After introducing his basic theory of mind and explaining its multi-modal nature, Damasio next explains the representational nature of brain imagery. Some images are the result of recalled memories or imaginings of the future, but the most important source of imagery are the early sensory cortices. Damasio explains the image-generating mechanism for sensation because recalled and imagined images are based on it. Memories and imaginings are known to activate the same "circuitry" as does sensation. The early sensory cortices are just the first parts of the brain to receive input from the sense organs.

As Damasio uses the terms, images are based on representations. The early sensory cortices form representations of sensation, and the rest of the brain operates on these representations to form images. Amazingly, the representations that form in the early sensory cortices are "topographically organized." This is to say that the representations in these cortices visually resemble the objects they represent (Damasio 98, 104). The initial representations of sensed objects somehow look like the objects. Damasio only cites evidence that visual-cortical representations are topographically represented, but he seems to think that this aspect of mental imagery applies to the other modalities as well. He does not go into any detail about how, say, sound-representations resemble the original sounds, but he clearly thinks that representations in all sense modalities are topographically analogous to their originals (Damasio 99).

The topographically-similar representations are a necessary condition for consciousness, but not a sufficient one. In order to give rise to consciousness, these representations have to be associated with something else. Specifically, these sensory representations must join up with the brain's representations of the body proper (the body

outside of the brain). When sensory representations occur simultaneously with representations of the body, subjective experience occurs (ibid.). While sensory representations are the foundation on which mind rests, memories and imagined things also have to join with body representations in order to enter consciousness. When Damasio introduces this concept, that consciousness is the correlation of representations, he makes sure to point out that the self is a neurological state, not a homunculus. He again takes the effort to attack what he understands about Cartesianism.

As Damasio uses the terms, images are the stuff of consciousness. Images are the subjective experience of representations in the brain; conscious images form when sense representations are connected to body representations at the instant of stimulus. Thus consciousness is the state of perceiving images. Having sketched his theory of mind, Damasio next explains that the representations in the mind are not *stored* as facsimiles of the originals, even though the sensory representations were topographically similar to their originals. Memories are not stored as facsimiles of the original objects, but as what he calls "dispositional representations." These are sets of instructions which tell the sensory cortices how to activate in order to recreate the image of something previously experienced (Damasio 101-102). Memory is reconstructive, not photostatic, and the brain contains sets of instructions for reactivating the sensory cortices in order to recreate the experience of the remembered sensation. Damasio's choice of term, "dispositional representation," is not entirely clear, but it seems to be based on the fact that these representations predispose the brain to act in certain ways, to fire in certain patterns.

In this theory, all knowledge is contained as dispositional representations. These are instructions which tell the neurons how to fire in order to recreate the sensation

originally experienced. Acquired knowledge is stored in evolutionarily recent structures, such as the cerebral cortex and various sub-cortical nuclei. Innate knowledge, that with which people are born, is stored in evolutionarily more ancient structures such as the hypothalamus, brain stem and limbic system (Damasio 104-105, 127-128). Acquired knowledge is used for reasoning, planning and creativity; innate knowledge is used for body regulation and often does not enter the conscious mind.

Images in every sensory modality are the primary content of thought (Damasio 106-108), images are based on representations that are topographically similar to sensed objects, and these images are stored as non-topographic dispositional representations. That is Damasio's theory of mind, which he spends the rest of the text explaining, and which he will presently use to explain the strange behavior of the Gage patients described in the first four chapters. In order to apply this theory, Damasio goes on in chapter six to explain how his theory accommodates biological regulation.

Certain biological processes are just a part of everyday life and these are encoded into the brain in more-or-less static patterns, which are fixed by the genes. Both emotions and feelings are manifestations of "drives and instincts," and these are directly based on biological need. Drives and instincts either cause a being to act in a certain way, or else create a physiological condition that inclines the organism to act in a certain way (Damasio 115). Instincts thus are the conscious experience of a biological need. The neuroscientist does not go into a great deal of detail about this point, but one can extrapolate a lot from the minimal treatment he gives it.

Some activities of biological regulation operate beneath the level of consciousness because they do not require action to intervene. Some of this regulation is controlled

within the brain, but does not enter consciousness at all, such as hormone levels in the bloodstream. But some biological regulation does require conscious intervention, as in the case of hunger. A drop in blood sugar activates a dispositional representation in the brain that causes the sensation of hunger, which is a feeling or a perception of the body's condition. The person then acts to correct this condition by eating. And so it is with some other instinctive behaviors—they are programmed into the brain and operate either beneath consciousness or else they encourage the subject to voluntarily act to alter some circumstance (Damasio 116). Perception of bodily state also allows the organism to begin classifying experiences as "good" or "bad," based on how they affect the body. Any circumstance that causes the body to thrive would be good, and anything that causes harm to the body would be felt as bad. This observation, that good and bad derive from the condition of the body, will play an important role in the final explanation Damasio proposes. He thus proposes that moral judgments arise directly from the condition of the body, but also that social systems evolve to guide and sculpt instinctive behavior. He proposes that social learning simply supplements the innate capacities to increase an organism's chance of surviving. Social goals, he believes, trace back to basic biological goals.

Damasio has thus outlined a theory of mind based on representations. Thought is composed of operations on those representations; bio-regulatory drives (instincts) enter the mind through other representations. Damasio has thus incorporated both thought and instinct into his theory of mind. With all of this said, Damasio goes on in chapter seven to incorporate emotions into his theory. He quotes William James to introduce the basic idea that feelings are intimately connected with the state of the body: "What kind of an

emotion of fear would be left if the feeling neither of quickened heart-beats nor of shallow breathing…were present…it is quite impossible for me to think" (qtd. 129). Damasio proposes that "specific pattern[s] of body reaction" are programmed into the brain as innate dispositional representations (Damasio 130-164). For example, the brain has an innate disposition to cause the body to tremble in dangerous circumstances. The brain contains both a representation of the danger, but also a representation of the body's response to it. And it is the *perception* or feeling of the body's condition that constitutes the emotion fear. Emotions are something which can be felt; they are the experience of the body's pre-programmed responses to certain stimuli. Feelings are perceptions of body states, and emotions affect body states. But not all feelings are emotional (Damasio 143-145).

Some other emotions are just pre-programmed bodily responses to stimulus. The perception of these pre-programmed responses are emotions and require little to no cognitive evaluation. Apparently, over the course of animal evolution, certain patterns of response were so common that they became incorporated into the genes of the smarter animals. The extent to which emotions are programmed into the genes might be considerable, as Damasio suggests certain physical features innately inspire fear (sharpness, largeness and such) (131).

Damasio smartly anticipates a shortcoming to this aspect of his theory. Some emotions seem to have a larger cognitive component and require more conscious contemplation than others. Thus, Damasio makes the distinction between primary and secondary emotions as soon as he introduces this definition of emotion. The primary emotions are happiness, sadness, anger, fear and disgust (Damasio 149). These are the

pre-programmed emotions (genetically determined) and depend on evolutionarily older circuits, including the amygdala, limbic system, and the anterior cingulate.

The secondary emotions are subtle variations on the primary ones. Secondary emotions occur when a stimulus or a thought activates certain dispositional representations, which in turn cause the body to respond in certain patterns (as with the primary emotions). The difference is that the secondary emotions depend on more recent circuitry than the primaries, particularly the frontal cortices and the adjacent prefrontal areas. These emotions are more sensitive to learning and experience than the primaries. Secondary emotions result when sensory input enters the prefrontal area and activates a separate set of dispositional representations than those which correspond to the primary emotions (Damasio 137-139). The secondary emotions can be activated by more subtle cognitive content (processed by the frontal lobes) and they involve more subtle physiological responses than the primaries. The perception of these body states are the secondary emotions proper. Euphoria and ecstasy are variations on happiness; melancholy and wistfuness are variations of sadness; panic and shyness of fear. Other feelings such as *Schadenfreude* and embarrassment are also variations on primary emotions (Damasio 149-151).

In addition to the feelings associated with the emotions, there are also important "background feelings," which are the moment-to-moment representations of the body in the brain. Anosognosics have defective mechanisms for generating these feelings (ibid.). These feelings enter into consciousness and take part in reasoning just as the emotions do. Anosognosics cannot reason correctly about their physical conditions because they are not receiving correct information about their bodies (Damasio 154).

In just the same way that background feelings enter the mind and participate in reason, so too do the feelings of emotion enter the mind to participate in reason. Emotions are feelings about the body (Damasio 159) and can affect decisions about the body—most decisions ultimately trace back to the body. In the theory that Damasio is proposing, emotions are "just as cognitive as any other perceptual image, and just as dependent on cerebral-cortex processing as any other image" (ibid.). Although Damasio is somewhat vague about the term "cognitive," he seems to mean that emotions are appropriate responses to stimuli, that are a form of data-processing, they are a form of decision-making themselves. They also contribute to reasoning proper (Damasio 159, 164-165). Emotions are not just random sensations which have nothing to do with rationality.

With emotion now in his theory, Damasio introduces in chapter eight the centerpiece of his new theory of mind, the somatic-marker hypothesis. His goal in this book is to show that reasoning consists not merely of using logical strategies to decide among available options, but that emotion somehow factors into the process (Damasio 166, 171). The SM hypothesis is that emotional responses, both primary and secondary, are perceptions of how various stimuli make the body react. That which generates an unpleasant response is bad, that which generates a pleasant response is good. What Damasio terms the "high reason" view belongs to Plato, Descartes and Kant, and holds that "formal logic will, by itself, get us to the best available solution for any problem" and "Rational processing must be unencumbered by passion" (Damasio 171). But logic alone does not seem to work for the Gage patients. Elliot could reason quite well, but not perform tasks which ostensibly depend solely on reason. So, Damasio proposes that

emotions are (cognitive) evaluations that rest on the state of the body, and these supplement logic. Somatic markers are feelings stored as dispositional representations which mark a situation as good or bad in reference to the body (thus "somatic marker") (Damasio 173-175). He described these markers as a sort of "biasing device" that enter into the decision-making process, filtering some information and options. They *presumably* highlight the salient features of a situation (Damasio never quite says this, but he does talk about framing, which relates to salience). Somatic markers help one to choose the best option, highlighting those options which are best and worst relative to the body. The best option, naturally, would be that which most likely leads to survival, pleasure, and happiness. The somatic markers sift options and allow the logical devices operant in the mind to pick the best options (Damasio 173-175, 179, 189). Thus, feelings participate in the more rational functions of consciousness. They speed up the decision-making process, allowing the brain to immediately rule out some options, and predisposing it toward others. They also increase the accuracy of the process, by pre-selecting the options most likely to lead to survival.

1.5 Conclusion

Damasio ends his book with four short chapters. Chapter nine discusses additional tests he ran to confirm the somatic marker hypothesis. Chapter ten discusses his version of an embodied mind hypothesis. Chapters eleven and twelve discuss broader philosophical implications of the book. As interesting as this material is, it does not add substantially to the content of the theory.

As to the Gage patients, when their prefrontal areas are damaged, their secondary emotions and somatic markers no longer function or do not function as well, depending

on the extent of the damage. The prefrontal area is apparently where some of the processing of secondary emotions occurs and where somatic markers do their work (Damasio 211). Gage patients can still feel primary emotions, whereas patients with damage to the amygdala and limbic system can feel neither primary nor secondary emotions. But prefrontal patients can no longer reason correctly over problems because their somatic markers are not reaching the primary emotional circuitry. Their emotions cannot affect their reasoning, thus their reasoning fails to operate.

Damasio has thus refuted all four of the Cartesian concepts he aims at. (1) The mind is clearly generated by the brain. (2) Social processing, ethical reasoning, and the "higher" features of mind are not the products of an immaterial mind, but are carried out in the prefrontal areas, and draw on other brain structures. (3) Reason depends on emotion, without it people cannot reason. (4) The body and brain are an integrated unit and should not be thought of separately. The brain evolved to tend to the body.

As stated before, the problem is not that Damasio is necessarily wrong in his central contentions. The problem is that his understanding of Descartes is incomplete. He does not quite seem to grasp that computationalism is an outgrowth of the very "high reason" tradition he opposes. If he had understood this, he might not have proposed a computational-representational theory of mind. While his treatment of Descartes is somewhat amateur, this is no severe indictment of Damasio's work. His primary thesis is correct: emotion does participate in reason. Nonetheless, Damasio makes an error here, but it is a fortuitously instructive one. Like Descartes before him, if someone had not made this mistake, then later thinkers would not be able to untangle it.

CHAPTER TWO: THE PROBLEM OF CONSCIOUSNESS

2.1 Introduction

Damasio's theory of mind raises any number of difficulties. Probably the most serious shortcoming of the theory is its explanation of consciousness. For current discussion, consciousness refers to subjective experience or first-person phenomenology. Damasio himself must have noticed that consciousness was the weak part of his theory, for he dedicated his next book, *The Feeling of What Happens* (1999), to shoring up just this part of the theory. According to the neuroscientist, consciousness results from the juxtaposition (temporal superimposition) of otherwise insensate representations. The brain represents the subject's body and the world beyond the body. At the same time that a representation of an object enters the brain, the brain also represents the body in the process of representing an object. The brain notices an external object, but also notices that it is observing a representation of that object. On the one hand, this is a very astute observation—conscious experience is not merely observing a thing, but knowing that observation is taking place. On the other hand, this is an unsatisfying explanation of subjectivity, it feels somehow incomplete.

It turns out that Damasio, despite his own claims to the contrary, makes a mistake very like what Descartes did in his homuncular theory of mind. In his language and explanation of consciousness, Damasio presupposes subjectivity on the part of unconscious entities. His explanation attributes to the brain a quality of mind, the ability to observe subjectively, when it should be explaining how that quality occurs in the first place. Even on close inspection of the theory, it turns out that Damasio is unaware of having done anything wrong. He begs the question of consciousness without realizing it.

Damasio's "homuncular" mistake scuttles his explanation of subjectivity, but his theory does still describe the operations of the brain which produce subjective experiences, including emotion and instinct. There are parts of the brain which seem to be maps or representations of the body, and the brain does process sensory data in a process which can rightly be called representation. And the phenomenology of emotion does agree with the superimposition process which Damasio describes. Thus, while the theory cannot explain *why* consciousness arises, it can still function as a description of the mechanical part of consciousness. Given the intractability of consciousness as a philosophical and scientific problem, this mechanical description may be the best anyone can realistically expect from a theory of mind.

2.2 The "Homuncular" Mistake

Although Damasio wants very much to escape Cartesian philosophy, he just does not manage it. He understands Cartesianism to include mind-body dualism and the homuncular-theatrical theory of mind, but does not understand the homuncular mistake in depth. Thus, he makes something similar to that same mistake. In *Consciousness Explained* (1991), Daniel Dennett explains how Descartes made the original homuncular mistake, and why later theorists have sometimes repeated it:

> Even the most sophisticated materialists today often forget that once Descartes' ghostly *res cogitans* is discarded, there is no longer a role for a centralized gateway, or indeed for any functional center to the brain. (106)

Damasio cites Dennett's objection to this Cartesian theater, but he only understands part of that objection. Damasio knows that there is no central viewing place in the brain where all the information in the brain comes together and is observed by an observer (Damasio 94). However, Dennett makes a larger point in his objection to

theatrical consciousness. The homuncular mistake occurs because of the presupposition of subjectivity on the part of the contents of the mind:

> The brain is Headquarters, the place where the ultimate observer is, but there is no reason to believe that the brain itself has any deeper headquarters, any inner sanctum…. In short, there is no observer inside the brain. (ibid.)

Dennett's objection is rather difficult to grasp, fortunately he spends an entire chapter on it. What he means is that there is no single point in the brain at which pre-experiential data become conscious information, no "mental divide" (Dennett 109). There is a "spatiotemporal smearing" of observation across much of the brain (Dennett 126). Observation does not occur at a specific place or time, it is not a localizable process. Subjectivity is essentially just the way people characterize the sum of mental processes; the appearance of observation is essentially an illusion. Damasio comes very close to this concept of mind by supposing that the theatrical nature of phenomenology, the sense of being an observer in a theater, is merely the result of the simultaneous function of different brain areas. Damasio comes very close to Dennett's idea, but just misses it.

For Dennett, attempting to locate subjectivity in some specific function of the brain is essentially the same as locating it in some specific area of the brain. In both cases, the theorist has to presuppose one part/process of the brain to be capable of magically generating subjectivity, of already possessing it. Damasio invokes the subjectivity which an observer would bring to the mind; he still talks about the mind as though there were already a subjective presence in there. The observer might not be located anywhere in the brain, but observation still takes place. To presuppose observation is to presuppose an observer. Damasio does precisely this. The

neuroscientist offers his own summary of his theory of consciousness, which warrants quoting at length:

> We become conscious […] when our organisms internally construct and internally exhibit a specific kind of wordless knowledge—that our organism has been changed by an object—and when such knowledge occurs along with the salient internal exhibit of an object. The simplest form in which this knowledge emerges is the feeling of knowing….
>
> Core consciousness occurs when the brain's representation devices generate an imaged, nonverbal account of how the organism's own state is affected by the organism's processing of an object….
>
> The organism, as a unit, is mapped in the organism's brain, within structures that regulate the organism's life and signal its internal state continuously; the object is also mapped within the brain, in the sensory and motor structures activated by the interaction of the organism with the object; both organism and object are mapped as neural patterns. (Damasio 168-169)

This passage is from *The Feeling of What Happens,* in which the neuroscientist focuses exclusively on his theory of consciousness, but this description does not disagree in any major details with the concepts presented in *Descartes' Error*. There, he makes the statement:

> Brains can have many intervening steps in the circuits mediating between stimulus and response, and still have no mind, if they do not meet an essential condition: the ability to display images internally and to order those images in a process called thought. (Damasio 89)

He later explains that "images are probably the main content of our thoughts, regardless of the sensory modality in which they are generated" (Damasio 107). He argues here that words and abstract symbols (such as mathematical or logical formulae) are not the primary content of thought, but that images are (Damasio 106-107).

The subjectivity mistake is present in these passages. Damasio inadvertently constructs a homuncular explanation of mind by talking about subjectivity as though it were a localizable process. The mistake here is to presuppose mind (subjectivity) on the part of mindless objects, and then claim that those mindless objects produce mind.

Essentially, this explanation just begs the question of consciousness. But because Damasio uses subjective rhetoric to explain mind, the resulting explanation sounds like a theory of consciousness, though it is not.

The quotes from *Descartes' Error* show this mistake directly. Thoughts are supposed to be made up of images, and mind results only when the brain forms images. Thus mind is essentially a series of images. But if there are images and the mind is conscious, then presumably there must also be an observer. This presupposed observer must be the source of consciousness, because there is nothing else in the theory to provide it. Either the images are themselves conscious, or else the juxtaposition of different images produces consciousness. Damasio favors the latter explanation, but does not explain how it happens, which is why he revisited the subject in the later book.

The explanation given in *The Feeling of What Happens* is supposed to clear up this issue, but does not. There Damasio posits that the brain forms representations of observed objects, but at the same time it forms detailed representations of the body in the act of observing. The coincidence of object images and self-observing-object images should produce consciousness. He says that a human organism becomes conscious when it "exhibits a specific kind of wordless knowledge." But the word "exhibit" sounds suspiciously like a conscious, deliberate act of display, and it also requires that some conscious entity be present to observe the exhibition. Thus, he seems to suggest that there is a conscious entity exhibiting knowledge and a conscious entity observing it. He also uses the word "knowledge," which contains an element of subjectivity as well. Unthinking, unconscious entities such as books and computers are able to represent knowledge, but do not themselves possess it. Knowledge is generally the word reserved

for the contents of a living mind, it is the word which names the subjective experience of containing a representation.

Further, the exhibition of knowledge enters consciousness as the "feeling of knowing." The term "feeling" does not refer to the kind of detection of which a mechanical device is capable, but to the experience a sentient being has when he encounters something. The sentient experience of feeling includes the ineluctable quality of the experience—the redness of red, the painfulness of pain, etc. Here, again, Damasio explains consciousness by attributing subjectivity to a brain which his theory says is unconscious before the process of representation.

This subjective rhetoric is interspersed among more appropriate mechanical language which does not presuppose subjectivity. The neuroscientist refers to "representations" and "neural patterns," neither of which is necessarily conscious. Representations are currently the dominant method of explaining cognition mechanically, and the neural patterns which interact to create consciousness are not themselves conscious. Damasio uses subjective and mechanical rhetoric together without explaining the connection between the two. Representations become knowledge and feelings, and neural patterns become sentient, subjective entities: "core consciousness occurs when the brain's representation devices generate an imaged, nonverbal account." Damasio's mistake is an unintentional, maybe even accidental, equating of subjective and mechanical rhetoric. He does not explain how the juxtaposition of two or more insensate representations, or the interaction of unconscious sets of neurons, can create a subjective experience. He just introduces the subjective rhetoric and acts as though this explains consciousness.

Attributing consciousness to unconscious entities, then explaining consciousness in terms of those entities begs the question of consciousness. Begging the question is a serious enough error, but even if Damasio somehow retooled the rhetoric in his theory, he would still not have an explanation of consciousness. If the implied observer were to fall out of the theory, then what is left is the interaction of insensate, unconscious objects. This just does not do the work which a proper theory of consciousness has to do. William Seager explains in *Theories of Consciousness* that such a theory has to attempt to solve the "generation problem," which is precisely the problem of how unfeeling matter generates subjective experience. In his own words:

> The generation problem can be vividly expressed as the simple question: what is it about matter that accounts for its ability to become conscious? We know, pretty well, how matter works, and there is no sign of consciousness in its fundamental operations […] nor in the laws by which matter combines into ever more complex chemical, biochemical, biological and ultimately human configurations. (Seager 18)

Seager surveys attempts to deal with this problem from Descartes to the present, including identity theories, panpsychism, representationalism, and various attempts to dissolve the problem. He concludes that none of the approaches produced by science or philosophy satisfactorily answers the generation problem.

According to Seager, representational theories of mind, of which Damasio's is an example, face the particular problem of explaining why only *some* of the many representations in the mind are conscious, while vastly more are not. The problem is that if representations generate consciousness, then all representations should be in consciousness, but this is clearly not the case. The example he cites is of human stereo vision. In normally-sighted people, both eyes generate simultaneously separate

representations of the same object. Each of these representations can independently enter consciousness, as when one eye is closed. But when both eyes are open and facing the same object, a single composite representation enters consciousness—the two independent representations lose their status as objects of consciousness. If each representation independently conferred consciousness on an observer, he would have double-consciousness of the same object, rather than consciousness of the composite (Seager 164-165). People with dramatically different levels of vision in their two eyes learn to be aware of only the stronger eye's representation—the other image almost completely drops out of consciousness. If representations generated consciousness, it is hard to understand how some representations could fail to be conscious.

Interestingly, Damasio's theory of mind does offer a partial explanation of this particular problem of representational consciousness. Damasio's theory posits that only those sensory representations which interact with representations of the body enter into consciousness. If something happens to the body and it does not enter consciousness, that is because the event was not imaged. If a representation forms in the brain but does not connect to body images, it does not enter consciousness. But if an image forms and does fire some of the neurons responsible for representing the body, then that image enters consciousness.

While Damasio's theory offers a possible solution to this one shortcoming of representational theories of consciousness, it fails in the more fundamental task of solving the generation problem. As Seager discusses it, consciousness seems to be separated from matter by an almost ontological divide—matter and mind appear to be two very different things conceptually. No attempts to explain mind in terms of matter

have yet succeeded. Indeed, it is not even obvious what a materialistic theory of consciousness would have to look like in order to satisfactorily solve the generation problem. Damasio's theory performs a novel maneuver when it suggests that the *interaction* of representations generates consciousness. And maybe this is the case, maybe the interaction of body-images and world-images is the source of consciousness. Maybe consciousness is just the phenomenal space between simultaneously held images, but Damasio does not explain how to bridge the quasi-ontological gap between mind and matter. He just does not offer anything which could explain mind at its most basic level.

Damasio insists repeatedly that his view of the mind is not homuncular. And it is not homuncular in a literal sense, there is no little man nor a pineal gland which magically confers consciousness onto the brain. Damasio is aware of the danger of making this mistake, but he does something very much like it anyway. He relocates the homunculus to functional space, which is to say that the observer is not a physical object, but observation is implicit in the process of representing or imaging the world. This illustrates the subtlety of the homuncular-theatrical mistake.

Still, one must wonder why a man as smart as Damasio would make a mistake which he is deliberately guarding against. This is because the theory still works as a description of the processes which underlie the conscious mind, even though it cannot explain consciousness itself. Damasio's mistake was trying to identify consciousness itself with the underlying representational process and mechanisms, the neurological substrate of conscious experience. At heart, argues Seager, this is the mistake which all identity theories make (Seager 47-48).

2.3 The Theory as Mechanical Description

Damasio's theory does serve at least as a mechanical description of consciousness because it describes the mechanisms of subjective experience and the explanation accords with the phenomenology of such experiences. He posits that emotion and instinct result from representations of the body. Emotions begin when the brain forms an image of an object and this image affects the body in the manners which correspond to anger, fear, happiness, sadness, disgust, etc. The emotion is complete when the brain images the body in the process of reacting emotionally. Instinct is when the brain's representation of the body displays an imbalance of some chemical factor. The brain then creates an urge to balance that factor. This theory posits the existence of object representations and body representations and requires that they interact. Damasio presents empirical evidence for the existence of both kinds of representations, thus all the remains is to justify terming them such.

Object representations are present in the form of memories. In order for a person to recognize a previously-viewed object, there has to be *some* sort of record of the object in the brain, and that pattern must somehow be isomorphic to the features of the original. Damasio explains memory as a two-stage process. Memories begin in "early sensory cortices," which are the parts of the brain that first receive signals from the sense organs. There are different cortices for different sense modalities. Damasio explains that these parts of the brain form "topographically organized representations" of the original objects (Damasio 98, note p. 104, fig. 5-2). This is to say that the brain forms a neural pattern in the likeness of the original, that the neural object resembles the original. These topographical representations are later stored as "dispositional representations," which

are not themselves topographically-organized. All knowledge, whether acquired or innate, is stored as these dispositional representations (Damasio 102-105). This type of representation is not an image of anything, but a stored pattern of neural firing. Each such representation is a set of operating instructions, which instruct different parts of the brain to fire in certain patterns. When someone recalls a memory, the dispositional representation for an object activates and instructs the early sensory cortices to reconstruct the image of the object as it first formed there (ibid.). Remembering something feels so much like seeing it again because memory activates some of the same "circuits" that activated during the original viewing. The dispositional representation is like a digital recording of a sound or image, which does not resemble the original, but can instruct a digital device to reconstruct the stored object. Thus memory is a process of representation, in the sense that a likeness of an object forms in the brain, and is then stored in another format.

The second sort of mental representations Damasio describes are of the body. There are substantial sections of the brain which clearly do function to monitor the body. The behavior of these "maps" corresponds to the behavior of the body and thus the maps are representations of body. The Penfield homunculus is the most notorious of such body images, but it is not the only one. There are somatosensory brain areas which pay particular attention to touch, temperature (as felt by the skin), pain, and the sensations generated by muscles, joints and viscera (Damasio 65). The somatosensory regions sense what is occurring in the body just as other areas sense what is happening outside the body.

A radical defense of Damasio would assert that distinguishing between sensation and body state is arbitrary. To see is to know that the eye (a part of the body) is receiving light. To feel heat is to know that some part of the body is receiving heat. To feel pain is for some part of the body to be damaged. Thus, sensation can also a case of body-representation, as are the sensations of joint position or pain.

In any case, somatosensory areas generate an "integrated map" (different maps and data are integrated) of the state of the body (Damasio 66). The right hemisphere seems to do more of the work of monitoring the body than the left, because damage to the right somatosensory cortices has a more profound effect than damage to the left. Anosognosia results from damage only to the right somatosensory regions. In this condition, patients are unaware of physical damage and handicap. Such a patient might be unable to move one of his arms, but does not know it unless told. The victim will be genuinely surprised when informed of his defect, and some will even forget about it later. These patients still think, reason and still sense the outer world, but their brains are no longer informing them of the correct condition of the body (Damasio 64-66). The infamous "phantom limb" syndrome occurs when certain brain areas responsible for monitoring the body are damaged, but the background static maps of the body are still intact. Without updated information about the state of the body, the victim feels the body as it is represented in the intact static maps, which are built into the brain by the genes. Both the anosognosic and the phantom limb victim perceive their bodies incorrectly because of how the bodies are represented by the brain. Sensation results from those representations of the body, not from the actual state of the body itself. And so it turns

out that Damasio is again correct to think that there are representations present in the brain.

Damasio's theory posits that these object and body representations interact to produce emotions. This interaction would be obvious from simple readings of brain scans. It is not likely that he would read the scans incorrectly, so it seems likely that the neurons which create body images and those which create object images fire at the same time. But it is not necessarily the case that the firings of these two sets of neurons are both part of emotion. However, the phenomenology of emotion agrees with Damasio's theory on this point. The five primary emotional states (happiness, sadness, anger, fear and disgust) do involve distinct body states. In this theory, an emotion is not just an empty representation like the word "anger." Each emotion is an experience, a thought which is felt because it is played out in the body. The body signals a change to the brain, which incorporates that change into the representation of the body. And the representation is felt as the experience of emotion. This is a fair description of the experience of emotion. And so, though Damasio's experience does not explain why things feel as they do, he does at least describe the process which underlies subjectivity.

2.4 Conclusion

Damasio admits in the first chapter of *The Feeling of What Happens* that he cannot explain qualia. Although the theory fails to explain subjective consciousness, it does succeed in describing some of the mechanical parts of mind. Damasio provides a plausible explanation of how memory, instinct, and emotion operate at the neurological level. Instinct and emotion depend on the representation of body states. These representations have a subjective feel about them, people respond to body states

according to the way they feel. But these subjective sensations arise from objects, from things which can be observed. And much of what Damasio proposes does accord with phenomenology, not to mention the available evidence of neuroscience. Emotion does involve intense body states; memory is like experiencing something again; and consciousness consists of sensation and the awareness of sensation. Just because the theory fails to explain why things feel as they do, it does not follow that his explanation of the mechanics is also mistaken. It is entirely reasonable to expect a mechanical description of consciousness to precede an explanation of qualitative consciousness. It is enough for a theory to observe that neural objects have a feeling about them and to explain how these neural objects behave.

CHAPTER THREE: DAMASIO, "GOFAI," AND EMBODIED COGNITION

3.1 Introduction

Through most of the twentieth century, the dominant paradigm of mechanistic theories of mind was computation. The underlying belief was that, like a Turing machine, the brain contains symbols and operates on these according to logical rules. This is the heart of any computational theory of mind (CTM). The rise of digital computers allowed researchers to build machines to test this belief, and they met with at least modest success. The collected attempts to emulate human intelligence with computers is what John Haugeland calls "GOFAI," for "Good Old Fashioned Artificial Intelligence" (Haugeland 112). It is "old fashioned" because new approaches to AI emerged in the wake of GOFAI's many failures. The major contemporary alternatives to GOFAI are connectionism and embodied cognition. The central tenet of connectionism is the belief that intelligence arises from the *patterns* of neurons in the brain, rather than from the processing of symbols. Embodied cognition (EC), in its most radical formulation, holds that a centralized symbol-crunching computer is not necessary to produce intelligent behavior. While connectionism is interesting in its own right, it is rather less radical than embodied cognition. A connectionist network is still a centralized computing device, one in which symbols are replaced by patterns of neurons and firing strengths (Clark, *Mindware*, 62). Thus, connectionism can be understood as a subset-of GOFAI in this context, as much as some theorists would object. Embodied cognition is a much more radical approach than connectionism because EC is decentralized and does not rely on symbol processing. The body is the processing unit and the world "is its own best representation" (Rodney Brooks, qtd. in Clark, *Being There*, 46).

On first glance, it is not apparent at how Damasio's theory of mind should be understood in relation to either GOFAI or EC. Damasio's theory is representational, like GOFAI, but its representations are of the body. The representations in Damasio's scheme are not the empty lingua-form symbols of a Turing machine, and the operations therein are not the strict logic of computers. Moreover, Damasio's theory uses the body as the core of computational cognition. The theory uses the body to think, but not in exactly the same way that EC does. Thence, it is not apparent where Damasio's theory fits into the large realm of cognitive science.

Cogburn and Megill (2005) suggest that Damasio's ideas can supplement GOFAI, that the somatic markers offer a possible solution to the "frame problem" of cognitive science. Coupling Damasio to GOFAI might solve the worst problems of computational theories of mind. However, GOFAI is such a deeply flawed approach that nothing might be able to save it. Coupling Damasio's theory with EC not only makes more sense conceptually, but appears to lead to a more powerful theory of mind.

3.2 Symbolic Computation and "GOFAI"

The modern digital computer originated conceptually with Alan Turing's hypothetical device. A Turing machine would consist of a sort of simple processor with onboard instructions (like a modern CPU), a reading/writing head, and a long line of tape with symbols imprinted on it. The processor has enough memory to recall what symbol was last read and what state the machine is in. Given instructions for performing some task like addition, the Turing machine would receive a series of symbols. It would then read the symbols, operate on them inside the processor, and rewrite the appropriate output to the tape. Input values have to be encoded symbolically according to the logic

of the system, and the answer has to be interpreted accordingly. Turing showed that such a device should be able to solve any sufficiently well-specified problem (Clark, *Mindware*, 10-12). This was the beginning of modern computation.

Shortly after Turing, digital computers became available and generations of programmers and engineers advanced this technology. A program written in the contemporary language C++ would compute just as does the Turing device sketched above. The technology and syntax are vastly more sophisticated than what Turing envisioned, but the principles are the same. A C++ program might read from a data file which looks something like this:

```
A    6    7    7    8    5    6    -1
B    3    4    5    5    2    -1
C    6    7    6    7    6    7    8    6    -1
…
```

The letters on each line are ID tags, indicating that record A is on the first line, B is on the second and so forth. The positive numbers on the lines might represent number of eggs a person ate in a given week, or number of cigarettes he smoked in an eight-hour period, or anything. The negative ones are “end of line” sentinels, which tells the program that the data for the record is over. C++ syntax is sophisticated enough to recognize the end-of-line character embedded in data files, but the technique of marking the end of a record is useful both for the amateur programmer and for the purposes of illustration.

A C++ program for averaging the number numbers here would have instructions to read the ID tag and store that to temporary memory. It would then read the integers, add them to the total, store the total in memory, and continue until it read the sentinel. Then it would write the final total to a new data file, indicating that the average number

for record A is 6.5 or 7 (depending on whether the answer is formatted for real numbers or integers), the average for B is 3.8 and so forth. After computing the total for each line, the program would move to the next line as long as there are letters to read. When there are no more letters, it would close the source file, commit whatever final instructions it has, and end the program. What should be apparent here is that the techniques for computation are more sophisticated and agile than what Turing described, but the function is the same in its underlying principles. The system reads symbols, performs logical operations on them according to an algorithm, and outputs an answer if the problem is sufficiently well specified. The point here is that modern computation is still computation in the sense that Turing envisioned. No matter how sophisticated a modern computer program may be, at bottom it just processes alphanumeric symbols and logical operators.

The symbols in GOFAI are not intrinsically meaningful. They only gain meaning from the larger context in which they are applied. The hypothetical C++ program described above would have instructions in it that make it output a line of text interspersed with the numerical results. So, the program would have instructions to output "Client 'ID' ate 'average' eggs per week for 'count' weeks," which would output "Client A ate 6.5 eggs per week for 6 weeks." The program would "know" how many weeks each record describes because it would have kept count of the number of times it performed the read-add instruction. Though the program would output such text and this would certainly seem to indicate a thoughtful response, of course this is just computation. These numbers might have meant anything.

Semantics might seem to be a stumbling point for GOFAI, but the general assumption has been that the meaning of the calculated values becomes apparent in the larger scheme of application. Cognition, as people experience it, is only the sum of a great many simultaneous calculations. The computational landscape of a mind is so rich that each computation gains meaning from this context. As Haugeland said, "if you take care of the syntax, the semantics will take care of itself" (qtd. in Clark, *Mindware,* 9). Haugeland summarizes things quite nicely:

> A GOFAI system has an inner playing field, on which inner tokens are arranged and manipulated by one or more inner players under the supervision of an inner referee. These inner token manipulations are interpreted as the thought processes by virtue of which the overall system manages to make sense and act intelligently. In other words, the overall intelligence is explained by analyzing the system into smaller (less intelligent) components, whose external symbolic moves and interactions are the larger system's internal reasonable cognitions [sic]. That's the paradigm of cognitive science. (Haugeland 117)

Rather, this *was* the paradigm of cognitive science for several decades. Connectionism rose to prominence during the 1980's, and embodied cognition appeared in the 1990's. In any case, Turing-style computation used to be the accepted explanation for human intelligence. While human-like intellect has never been achieved by a symbol-manipulator, modern computers certainly can accomplish a great many tasks with the techniques of computation. Any sort of analysis which reduces to numbers and mathematical procedures, for example, are perfect tasks for GOFAI. But, as Hubert Dreyfus explained over thirty years ago, this technique is just not sufficient for producing genuine intelligence.

3.3 The Limits of GOFAI

Hubert Dreyfus wrote the major critique of artificial reason, *What Computer Can't Do* (later updated and released as *What Computers Still Can't Do*). Dreyfus identifies four assumptions that GOFAI or any computational theory of mind (CTM) has to make. He gives them in this order:

(1) The *biological assumption* holds that the *brain* must act like a computer or other device, performing mechanical operations on internally-represented facts according to internally-represented rules.

(2) The *psychological assumption* holds that the *mind* must also act like a computer, which is to say that thoughts (at the personal or sub-personal levels) must be analogous to the algorithms in computer programs.

(3) The *epistemological assumption* holds that (a) human knowledge can be represented by propositions or rules and (b) those propositions can be used to reproduce human behavior.

(4) The *ontological assumption* holds that the world of human experience consists of discrete and determinate elements.

Of these assumptions, (1) and (4) are the most defensible, though their veracity is still in question. The biological assumption has not yet been proven, but also not yet refuted.. While neuroscience has learned a great deal since Dreyfus about the workings of the brain, it is by no means obvious even now that the brain-computer analogy is accurate. Similarly, the ontological assumption is an uncertain claim. However, these two assumptions are not the real problems for GOFAI.

The real snags for any CTM come from the second and third assumptions. People do not seem to think quite like computers, as in the psychological assumption. And it is not obvious that human behavior can be described by rules, much less reproduced by them, as in the epistemological assumption. Computers contain representations of rules and information, and they perform logical operations (computations) on these representations. Dreyfus provides examples from the history of digital computers to show how the computational approach fails to reproduce human intelligence.

When Dreyfus discusses the psychological assumption, he identifies some major differences between human intelligence and mechanical calculation. The most conspicuous difference he identifies is what he calls "zeroing in," which is especially evident in the different approaches men and computers take to playing chess. According to Dreyfus, computers calculate tens of thousands of possible moves based on the configuration of pieces on the board. The computer then selects the best of these moves according to its set of chess-playing instructions. Working from the same board, from the same initial conditions, expert human players only contemplate tens or hundreds of moves (Dreyfus 102-103). Human players somehow just know what moves are the correct ones, without evaluating *all* of the possibilities. They are somehow able to zero in on the set of correct moves out of the vast set of possible moves. Whatever it is that people do when playing chess, it seems to differ from the "brute force" method that computers use (ibid.). In Dreyfus' words, "[T]hinking and perception involve global processes which cannot be understood in terms of a sequence or even a parallel set of discrete operations" (163). "Global" here refers only to the strange ability of the mind to deal with large sets of information as wholes, rather than as pieces. A person

demonstrates “global” thinking when he zeros in on a few correct moves out of the countless thousands of possible ones. A computer necessarily deals with the same situation piecemeal.

This zeroing in phenomenon shows a difference between men and machines that is very hard to explain away. The computationalist will insist that people must make thousands of calculations in order to play chess. If these calculations do not occur consciously, then they must occur subconsciously. The problem is that even subconscious calculations do not explain zeroing in. Dreyfus’ point is that zeroing in is not an algorithmic, stepwise process. A computer has to consider each possible move as if it were the right one, it does not have the immediate instinct that some moves are better than others. Thus, even the subconscious stratagem does not save the psychological assumption—human thought appears to be different from the algorithms inside a computer's memory. However, even if the Computationalist finds some way to dodge this problem, he still has to deal with the epistemological assumption.

The truth of the epistemological assumption speaks directly to the plausibility of any CTM. This assumption consists of two claims: (a) human knowledge can be represented by propositions and (b) those propositions can be used to reproduce human behavior. Dreyfus offers a stronger argument against the second claim. He points out that if a computer were programmed with a body of rules about language, the computer would have a hard time dealing with odd or incorrect utterances. Such a computer can only recognize an utterance to be correct or incorrect according to its rules.

A person can understand the meaning of a grammatically-incorrect statement, sometimes quite easily. Minor slips in grammar do not necessarily prevent

communication. For a computer, however, a grammatically-incorrect statement is meaningless. The computer would also have a hard time dealing with odd statements such as a non-literal utterance like “The idea is in the pen” (Dreyfus 198-199). A computer would have to analyze the terms “idea” and “pen” and conclude that this statement is not possible. The problem is that the rule-governed behavior of computers winds up being very inflexible, non-adaptive, unintelligent.

According to Dreyfus, this sort of inflexibility recurs over and over in Artificial Intelligence work. Researchers write a program to perform a specific complex task, one which closely mimics some human capacity. The program often succeeds impressively in its set task. But when the programmers attempt to apply their software to other problems, when they attempt to generalize to a domain of human performance, the program suddenly comes up short. Software, for example, which can identify geometric shapes suddenly becomes quite confused when light and shadow are introduced (Dreyfus 15-20). And simple acts of locomotion have been notoriously difficult for GOFAI robots to reproduce. Traditional walking robots seem more adept at falling down than anything else.

Faced with the shortcomings of their machines, programmers might try to solve the problem by adding more rules and more information to an expert system. But this does not give the computer a human level of competence. A computer would either need a near-infinite set of rules to cover every specific conceivable circumstance, or else it would need rules to govern the use of its rules, and rules to govern the use of these rules and so forth. In either case, the set of rules necessary to mimic human competence would have to be impracticably immense (Dreyfus 199-200). The point is that finite sets of data

do not seem able to generate the kind of creative performance, the adaptive competence, which humans manifest in their actions. Thus, the rule-driven approach to reproducing human performance seems incorrect. As Dreyfus says, "the general laws of competence cannot be directly applied to simulate behavior" (202).

GOFAI would seem to face an immense problem if the epistemological assumption is incorrect. Human behavior might not reduce to rules. If Dreyfus is correct, GOFAI is simply a wrong approach. Rules and propositions seem unable to reproduce human mentality.

3.4 Embodied Cognition

Symbolic computation can perform specific actions of minimal scope which appear to be intelligent, but which simply fail to produce intelligence in the realm beyond well-specified logical problems. Because computation fails in many real-world scenarios, some robotics researchers developed a technique which is about as far away from GOFAI as seems possible. They developed robots which do not have a central processing unit, thus their behavior is not controlled by a digital computer. These robots do not look at the world for symbols and attempt to apply instructions to those symbols. They also do not have the immense data "models" which are necessary to make computational systems operate in complex problems (Clark, *Being There*, 21-22). These models are the sets of data and instructions which are meant to guide the robot in any conceivable circumstance, and allow it to reason its way through unforeseen problems. However, such data models can grow improbably immense and still not produce adaptive behavior. Instead of using such models, these new robots use their bodies to interact with the world. They solve problems not with symbolic manipulation, but by trial-and-error

manipulations of their own bodies. This sort of embodied cognition (EC) has achieved surprising, even revolutionary, degrees of success.

A good example of embodied cognition are two six-legged robots built at Case Western Reserve in the early 1990's. Each of the six legs has a control circuit which can raise the leg, swing it forward, or swing it backward. There are also "inhibitory linkages," which are feedback sensors to let the legs "know" when the others are resting stable on the ground. The legs move in patterns in which one tripod of legs is stable while the other moves. Then the stationary tripod begins its own motion cycle (Clark *Being There*, 15-17). The legs move in sequence rather like the firing pattern of the spark plugs in *older* cars with distributor caps—the circuitry is built to fire in a certain sequence without instruction from a computer. This approach is more successful than GOFAI at moving robots around, but it also does this without an enormous effort of computation.

The differences between GOFAI and EC are considerable, but the distinction between the two is actually not as sharp as one might expect. The six-legged robots described above are clearly EC. On the other hand, a robot which visually surveys its environment and compares what it sees to the list of characteristics labeled "stable purchase" would clearly be GOFAI. However, intermediate cases are conceivable. Installing a Turing-style device to control the legs of each of the hexapods would not necessarily move it from the category of EC to GOFAI and would likely not make it as clumsy as a GOFAI robot (this is a project conceived by Rodney Brooks at MIT and described in Clark, *Being There*, 15).

Clark believes that EC is powerful because it is (1) decentralized and distributed across the robot's body and (2) does not use the immense data models which actually seem to slow down GOFAI (*Being There* 21-22). Of these two, it seems that the first is less important to EC than the second. Clark does not offer an *a priori* reason why locating all of the circuits in one centralized location would necessarily impact the machine's function. The hexapod legs are controlled by simple circuits—it just does not matter whether that circuit is located on the leg or in the thorax. Circuits in the thorax would simply require longer lengths of wire to connect them to the legs than circuits located directly in the legs. The only obvious difference centralization would make is that the centrally-controlled robot would not be as robust as the distributed version. Damage to the central control apparatus could paralyze an entire robot, but the distributed robot would still be able to move if one of its leg controls were damaged. However, in terms of function, Clark does not establish that centralization of control necessarily impedes a GOFAI.

More important is the second issue, the use of data models in traditional AI. This seems to be the defining characteristic of GOFAI. At heart, a traditional AI searches its environment for elements which correspond to a symbol in its program. This is at least loosely analogous to the function of a Turing device; for a GOFAI system, the entire world is like a line of symbol-bearing tape. If and when it locates the appropriate symbolic elements, it attempts to apply (lingua-form) rules for behavior to the situation. It may have to consult other rules to instruct it how to "reason" over this situation. And if Dreyfus's description of the history of AI is correct, this technique almost invariably fails when applied to real-world scenarios. EC is more successful than GOFAI in dealing with

some robotics problems, but it also qualitatively resembles what living creatures do. EC proponents claim that the hexapods move in some ways like the insects they resemble. As Clark says, EC "has all the flavor of real, biological intelligence" (*Being There*, 17).

EC works for simple robots and simple locomotion, but there is evidence that sophisticated animal brains operate like an EC, at least some of the time. An example is how human infants learn to walk. They try to use their legs at birth, but stop around two months when the legs get too heavy relative to muscle mass. Once the muscle mass increases, they start trying to walk again around ten months of age. The entire process of learning to use the legs seems to be controlled by the structure of the legs themselves, by the tendency of the legs to move in certain ways. The child seems to have to learn how to use its own body, and seems not to have an onboard blueprint or program to control the process (Clark, *Being There*, 40-42).

Another example of people acting like an EC is when they play the game Tetris, in which falling blocks of different shapes have to be arranged to make solid lines. What is interesting is that people playing the game will rotate the pieces several times before dropping them, in an attempt to "see" what the best configuration of pieces would be. This kind of transformation would be easy for a symbolic computer, but human beings change their environment rather than the contents of their own minds in order to make a computational problem easier. According to Clark, a sign of an embodied mind is when it manipulates the world in order to make thinking easier (*Being There*, 65-66). Tetris is an example, but so too are more mundane examples like moving one piece of furniture to determine what a new configuration of an entire office would look like.

The examples Clark cites of EC-like human behavior could be the result of computation; his evidence is not entirely convincing on the point. But there is a body of data to indicate that even the sophisticated human brain has to interact with the world and adjust the body in order to solve problems, like a simpler EC system. The human brain also resorts to simplifying the world in order to solve spatial transformation problems that should be easy for computation. This evidence suggests that the human brain does not always compute answers, that it behaves like an EC some of the time.

Embodied cognition has its advantages, but even strong proponents like Clark do not think it can entirely reproduce human intelligence. He warns that abandoning traditional AI techniques wholesale in favor of EC would be tantamount to washing away babies with "floods of bathwater" (*Being There*, 22). There is just too much about human intellect that seems computational or representational. Damasio actually explains the appeal of centralized, computational-representational theories of mind when he describes brains as intermediate to stimulus and response (*Error*, 89). Brains enable complex responses, as though sensory input is being manipulated in the manner that a computer operates on representations. It is hard to imagine that an EC would ever be able to use language, contemplate a complex situation, or imagine possible outcomes. Such a device would never have the experience of thinking about anything. There must be more to intelligence than embodied cognition.

3.5 Damasio I and II

EC and GOFAI have both proven to be insufficient paradigms of cognition, but it is not clear how they should be altered. The two approaches are so antithetical that they seem incapable of rapprochement, so combining them seems out of the question.

Damasio's theory might be able to function as a stand-alone theory of mind, but it is far too underdeveloped for that purpose. Alternatively, Damasio's theory could be coupled with one of the other paradigms to produce a stronger approach to cognition. Ultimately, pairing Damasio with EC makes more sense than joining his theory with GOFAI.

Damasio I would be the addition of Damasio's theory to GOFAI. Cogburn and Megill propose that Damasio's theory would be able to solve the infamous frame problem of cognitive science. This is the problem of determining salience. Whenever an autonomous robot is given a task, it has to process a vast data model in order to calculate an answer. The problem is that such systems usually tend to fail to identify the important elements of a situation, the rule-driven approach fails and the robot simply does not know what to do.

Cogburn and Megill propose that Damasio's frontal-lobe patients are suffering from a human version of the frame problem. They propose that the somatic markers assign sufficient priority to the elements of a situation that the computer can calculate solutions. Positive emotions make some representation of an object more desirable, negative emotions make one less desirable. The pleasantness or unpleasantness can affect how mind operates on the object. Emotions, which create somatic markers, essentially edit data to allow computational processes to occur (Megill 309-312). In the absence of emotional content (the pre-frontal cases), cognition fails. Such is the case of Damasio's patient, Elliot.

Damasio I is really quite clever, but it does not get away from the deeper problems of GOFAI. All four of the assumptions of AI which Dreyfus identified will plague Damasio I. On the other hand, Damasio II, the joining of Damasio with EC,

avoids all of those assumptions. Further, it seems more plausible that evolution would have designed a somatic mind (a mind based on somatic computation) to solve the problems a body would face rather than the more abstract problems a GOFAI is programmed to solve.

Damasio II seems more plausible than Damasio I because the ideas integrate more closely, and DII should be immune to the frame problem like DI. Under DII, an organism would have its body as something of the "front line" of cognition, but there would be a whole second tier of cognitive capacities supporting the body. Over time, the use of the body would become programmed in the brain, the patterns of motion would come to be represented in the brain. And the brain would learn to operate on the representations of body state (such operations are posited by Damasio's theory). Once these representations are programmed or learned, the system learns how to manipulate the body, but it can also modify those representations in the process of imagination. Clark's discussion of an infant learning to walk by feeling out its own body might be an example of embodiment programming representative cognition (*Being There*, 40-42). A DII system also matches up with Damasio's observation that the maps of the body in the brain are constantly updated with fresh information from the body. Damasio's theory seems to match up with EC better than GOFAI. Further, a DII system would be at least as computationally robust as the DI system proposed by Cogburn and Megill, but it might be even more robust.

An Embodied Cognition device would be utterly immune to the frame problem. If one of the Case Western hexapods could not find stable purchase, it would slide and stumble across a surface until it found such. An EC never faces the frame problem

because it is not calculating anything; bodies do not suffer framing problems. DI might be able to solve the frame problem, but DII would avoid it entirely because of the connection of the body to representational cognition. Whenever the calculations in the representational part of a DII system fail, control shunts back to the body, which can then feel its way through a situation. This is decidedly reminiscent of what happens when a person faces some new situation; whether the new situation require a physical or intellectual skill, people have to feel their way through it, even if they have guidance. A GOFAI cannot typically learn anything, but has to be programmed with its skills; DI system would be similarly limited. A DII system, on the other hand, would have the body present as the tool for studying the environment. The body and its somatic logic would provide a finite set of possible approaches to any possible problem; all one has to do is to feel out a situation, try different alternatives, and find what works. Embodied Cognition and a DII system should both be immune to the frame problem, whereas DI systems have to solve it.

The process of learning described above certainly seems to happen for physical tasks such as playing sports or learning to play a musical instrument, but something about the process is reminiscent of the way people learn complex intellectual skills. Language, chess moves, ethical conduct, and even mathematical calculations can all be represented in terms of body states. A sound results from a certain state of the vocal chords—and even expert readers sometimes move their lips when they read, because words are made up of body states. Chess moves are rule-governed motions through physical space, easily represented by images of body states. Ethical conduct rests on imagining the states of

other bodies. And mathematical calculations begin with the base-ten system of the fingers. All of this is far from conclusive, but it will make for extraordinary future work.

3.6 A Research Program for the Future

Damasio's posits that neural representations of the body are essential to cognition, that the body is the foundation and subject of most cognitive processes. If true, his theory could be an important complement to the Embodied Cognition paradigm of cognitive science. A machine or organism which uses somatic computations to supplement its embodied cognition (or vice versa) could potentially be more intelligent that either a GOFAI or EC. However, both empirical and philosophical research will be required to develop a combined theory of Somatic Computation and Embodied Cognition (SCEC). The first challenge for this paradigm would be to identify the neural representations of the body and to describe their behavior as representations. In order for SCEC to operate, there must be "somatic logic" built into the brain. This is to say that the body must be represented in a way that its representations can be manipulated in a quasi-computational manner. The "virtual body" necessary for somatic computation needs to be identified as a representation, and the brain's manipulations of this representation need to be treated as computational processes.

If the brain represents the body, it either does so with traditional data "models" consisting of rules and lingua-form symbols, or it does so by some other means. The rule-driven, data-modeling approach fails to generate intelligence, so something else must be happening. Damasio is never explicit about what kind of body-representation occurs in the brain. He mentions that there are groups of neurons in the brain dedicated to monitoring the body. Further, the body and brain communicate in great detail to each

other with both nerves and chemicals in the bloodstream. Thus, the maps of the body in the brain are constantly updated with new information. That is as far as Damasio goes with explaining the representation of the body.

However the body is represented in the brain, that representation cannot be of the lingua-form variety. Instead, it seems to be necessary that the brain contains a neurological object which somehow acts like the body. The rules of the body must be stored directly as patterns of neurons which mimic the body's action. The following is speculation well beyond what Damasio offers, but the previous discussion of centralization in EC suggests a starting point for the study of the virtual body. If a robot or organism were controlled by distributed, decentralized circuits, there seems to be no reason in principle why the control circuits could not be moved to a central location. The resulting circuit-board would have features which correspond to all of the functions of the body which the board can control. One circuit would connect to each part of the body which is directly controlled. This circuit-board is potentially an analogue to the body. The board would likely have fewer degrees of freedom than the body it controls; some features of the body would doubtless not be controlled directly by the central device. However, the control board might still be enough like the body to act as a representation of it. If the structural relationships present within the body (the effects each body part has on other parts) were present in this circuit board, then this neurological object could be the foundation on which a representation of the body is built. In the sense that this hypothetical circuit-board would correspond to the functions of the body, it could be the neurological object needed for SCEC.

This circuit-board would be connected to the body it controls directly, but could also be connected to some sort of computational "scratch space" in the brain. If the control circuits connect to neurons which are not connected directly to the body, these other neurons could be a computational area, the inner playing field Haugeland describes for GOFAI. Motor neurons certainly are connected to non-motor neurons in the brain; each neuron, on average, is connected to thousands of others. Thus the body's own control circuitry could be the foundation for a "virtual body." The features of this control board or virtual body could act like a set of rules for body-based computation—they are not linguistic rules, but somatic ones. These rules would be equivalent to "the legs move backward and forward," "the arms move up, down, left and right" and so forth. However, the rules would not be linguistic artifacts as here, but stored patterns of neuronal firing which correspond to the motions of the body. These patterns of body states/motions could be the representations on which a computational mind rests, rather than lingua-form rules of GOFAI. This is to say that the very neurons which control the body could also serve as the neurological object which SCEC requires. This is a promising area for future research into SCEC.

3.7 Conclusion

While the circuit board is entirely a point of speculation, SCEC is still promising in another respect. This paradigm would posit a mind which matches phenomenology much more closely than GOFAI. A somatic mind would operate on representations of the body's states and functions. This means that the mind would contain not lingua-form rules and representations of information, but the very patterns of body operation necessary to act in the world. Particular motions each correspond to a set of fired

neurons, and these patterns of firing (possibly Damasio's "dispositional representations") would be the subject of computation. Moving the arm in a circle is a describable set of neuron firings, and that set of firings can be manipulated like any other representation. The somatic mind would be able to remember past body states and past motions, but would also be able to imagine new body states and new motions. Thus, memories and imaginings would consist of concrete body states rather than the abstract content which lingua-form representations are supposed to embody. Human experience seems to consist largely of sensations, of the feeling of what happens. These are things which can be described by language, but cannot be computationally reproduced by language. A mind consisting of body states seems closer to phenomenology than any GOFAI could ever get.

Damasio's theory and SCEC might never be complete theories of mind, because neither seems able to explain subjective consciousness. Damasio thinks that the actions of body-representations can generate consciousness. He also thinks that consciousness comes only from representations and not from the body itself (thence "phantom limb" syndrome). Just as a point of philosophical speculation, this theory of mind might get a step closer to subjectivity by positing consciousness to be the correlation of representations with their objects. This would probably not get to consciousness, but might get a little closer to it, conceptually speaking. Nonetheless, even if a somatic theory of mind cannot explain consciousness itself, it could still be a better theory than anything else out there. GOFAI just does not work to reproduce human-style intelligence. If Hubert Dreyfus, Andy Clark and Rodney Brooks are right, it will never succeed in producing an authentic mind. However, embodied cognition by itself almost

certainly cannot get to the human mind either. EC cannot explain the computationally-tractable aspects of the mind such as language and imagination. SCEC would combine the best parts of both paradigms into a theory that might explain the phenomenology of embodiment while also satisfying the instinct that some sort of representations are present in the brain.

WORKS CITED

Clark, Andy. *Being There: Putting Brain, Body and World Together Again*. Cambridge: The MIT Press, 1997.

---. *Mindware: An Introduction to the Philosophy of Cognitive Science*. New York: Oxford UP, 2001.

Damasio, Antonio. *Descartes' Error: Emotion, Reason and the Human Brain*. New York: Penguin Books, 1994.

---. *The Feeling of What Happens: Body and Emotions in the Making of Consciousness*. New York: Harcourt, 1999.

Dennett, Daniel C. *Consciousness Explained*. New York: Little, Brown and Company, 1991.

Dreyfus, Hubert. *What Computers Still Can't Do: A Critique of Artificial Reason*. Cambridge: The MIT Press, 1992.

Haugeland, John. *Artificial Intelligence: The Very Idea*. Cambridge: The MIT Press, 1985.

Megill, Jason, and Jon Cogburn. "Easy's Getting Harder all the Time: The Computational Theory and Affective States." *Ratio (New Series)* XVIII 3 (September 2005): 306-316.

Seager, William. *Theories of Consciousness: An Introduction and Assessment*. New York: Routledge, 1999.

VITA

Christopher D. Pope is a native of central Mississippi. He was born in 1976 and holds the Bachelor of Arts (2000) and Master of Arts (2002) in English literature from Mississippi State University. After working as a clothier and high school teacher, he enrolled at Louisiana State University to pursue a master's degree in philosophy, which will be conferred in August 2007. He currently teaches at Baton Rouge Community College and will begin work on a doctorate in cognitive science in 2007 at the University of Louisiana at Lafayette.