Descartes, the Cogito, and the Mind-Body Problem in the Context of Modern Neuroscience

Author: Willam J. Hendriksen

Descartes, the Cogito, and the Mind-Body Problem in the Context of Modern Neuroscience

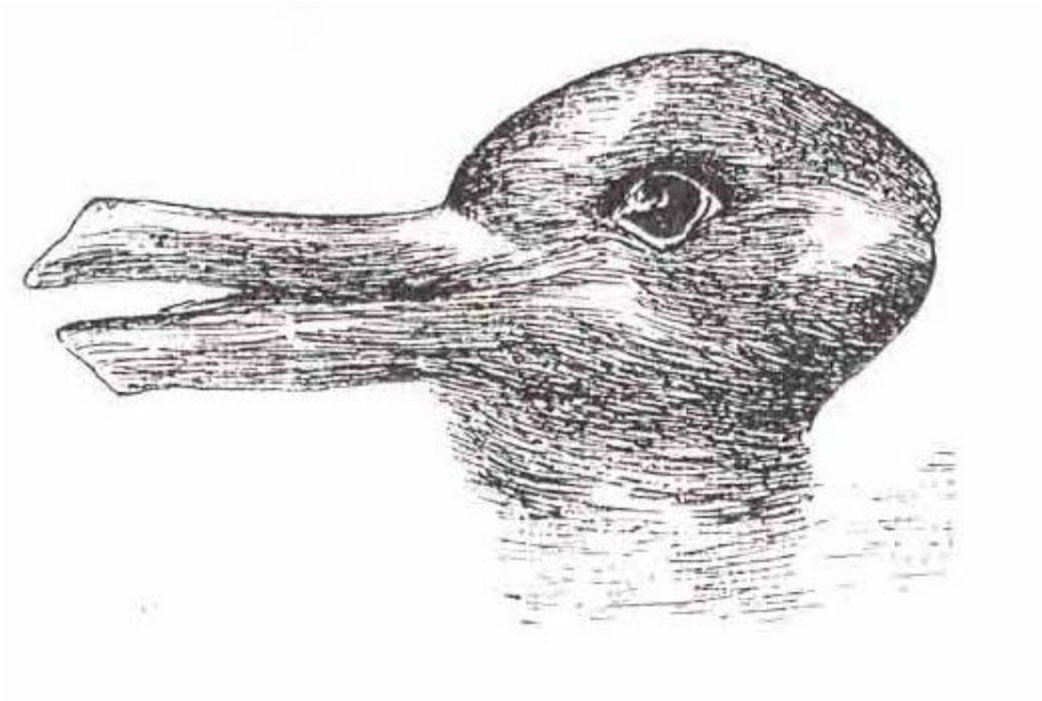


Image courtesy of Wikimedia Commons

William Hendriksen
advised by Marilee Ogren
Boston College
2009

Abstract:

The suggestion of a mind-brain duality that emerges out of Descartes' *cogito* argument is assessed in the context of twenty-first century neuroscience. The Cartesian texts are explored in order to qualify the extent to which the *cogito* necessitates such dualism and the functions that Descartes attributes to a non-corporeal soul are precisely defined. The relationship between the mind and brain is explored in the context of a number neuroscientific phenomena, including sensory perception, blindsight, amusia, phantom limb syndrome, frontal lobe lesions, and the neurodevelopmental disorder Williams syndrome, with an attempt to illuminate the physiological basis for each. Juxtaposing the two perspectives, the author concludes that Descartes hypothesis of a disembodied soul is no longer necessary and that a purely physiological understanding of the human mind is now possible, and that there is an underlying affinity between this assertion and Descartes theory of mind.

Introduction:

Descartes is perhaps most famously known for his statement, "*cogito ergo sum*," or "I think, therefore I am," and the notion of a distinction between the physical body and the mind, or disembodied soul that apparently arises from the *cogito*. This seems fundamentally at odds with the field of modern neuroscience, which seeks to explain human mental life in terms of neurons and their electrical activity. The present investigation seeks to counter that claim and demonstrate that rather than the impediment to a physiological understanding of the human mind that Cartesian philosophy has often been touted as, the two actually have a natural affinity and that substance dualism is not in fact an essential tenet of Descartes, but rather a largely anachronistic addition.

The study begins with a modern reformulation of the primary Cartesian texts with the intent of separating the distortions that have accumulated over the past four centuries from pure Descartes. What emerges is a less contentious system than what has come to be known as Cartesian dualism, and while Descartes is not absolved of hypothesizing a fundamental separation of mind and body and a disembodied soul, the author maintains that he does so as consequence of the ideological, scientific, and theological landscape of the seventeenth century. Further, the author asserts that such a hypothesis is no longer

necessary, and that it served as a placeholder in Descartes' theory of mind, a black box whose contents could not be understood within the confines of his environment, but one that we can now open.

The final assertion is that Descartes would readily concede to a purely neurophysiological explanation of the human mind and all of its capacities in the context of our twenty-first century understanding of these phenomena, that our present landscape and the conclusions that have emerged are not at odds with Descartes, but rather an increase in precision and refinement of theory that comes with the increased clarity that accompanies the passage of time in the evolution of scientific thought. The harmony between the essential tenets of Descartes, a desire for firm conclusions and increased explanatory power, and the modern quest to understand the physiological mind manifests when one removes these prejudices and the apparent dissonance evaporates.

The Mind-Body Problem Articulated:

René Descartes was a prolific writer who over the fifty-four years of his life contributed a wealth of theory and knowledge to both science and philosophy. This investigation makes no claim to examine the entirety of the Cartesian body of work, but rather to distill the fundamental ideas from his major works into a concise but representative statement of his philosophy as it pertains to the mind-body duality that emerges as an apparent consequence of the *cogito* argument. This task requires carefully navigating between the Scylla of simply choosing the elements of Descartes that appear most pertinent to the present discussion and failing to consider the larger context of his complete ideology, inevitably misrepresenting him, and the Charybdis of attempting to simultaneously consider all of Descartes' writing, which spans thirty years and is at times self-contradictory. Moreover the latter pitfall carries the further danger of including obsolete and outdated ideas that would render any serious attempt at criticism almost farcical. The final caveat is that, while Descartes is obviously the focal point of the investigation, the ideas take precedence over the man, and the mind-body problem that the author intends to submit to a neuroscientific investigation may not wholly represent the one realized by the man who first articulated it, though every attempt to minimize this drift is made, and the distortion of Descartes' ideas that has occurred over the past four hundred years is not what is intended by this concession. Thus the initial goal is to

consider the development of the *cogito* argument and its corollary of a dichotomy between body and mind as it is traced from Descartes' early writings, *Regulae ad directionem ingenii* (Rules for the Direction of the Mind, 1621, published posthumously) and *Le Monde* (The World, never published) and his philosophical treatises *Discourse on Method* (1637), *Meditations on First Philosophy* (1641), *Principles of Philosophy* (1644) and *Passions of the Soul* (1649), and to extract from this body of work a composite statement of the mind-body duality as Descartes would have framed it. The *cogito* itself only appears explicitly in *Discourse on Method* and *Principles of Philosophy*, and Cartesian dualism today is quite precisely and homogeneously defined, but rather than accept these principles at face value, the author intends to begin the present investigation as Descartes does his and "raze everything to the ground and begin again from the original foundations[17]."

Descartes' body of work reveals a characteristic shift from the overtly scientific and mathematical style of the *Regulae* and *Le Monde*, his earlier unpublished works, to the philosophy, theology, and metaphysics that characterize his subsequent writing. In *Le Monde* he explores physiology, astronomy, and physics. Another motif that runs through nearly all of his works is the question of epistemology. The *Regulae*, though never finished, was intended to consist of thirty-six rules divided into three groups of twelve, the first of which was concerned with basic principles of properly ordering knowledge, the second with addressing problems where the question could at least be framed and understood irrespective of whether a solution was available, and the third where the problem itself could not be perfectly understood[39]. This epistemological question continues through *Discourse* and *Meditations* where Descartes introduces the concept of radical doubt and abandons deductive reasoning based on sensory observation in search of a first principle of which he can be absolutely convinced. From this emerges the *cogito*, "I think, therefore I am," which he articulates explicitly in *Discourse* and *Principles* and alludes to in *Meditations*. Having established that *he* exists, and identifying himself as a "thinking thing," or *res cogitans*, he proceeds to surmise whether in fact his body exists, and from this emerges the conclusion that mind and body are two different types of substances, with two different essential qualities – thought and extension, respectively, and are therefore necessarily separate. This argument, more or

less, though with slightly different emphasis, unfolds in each Descartes' three major philosophical works – *Discourse, Meditations,* and *Principles*. The full title of the intermediate work, *Meditations on First Philosophy in which the Existence of God and the Distinction between the Soul and the Body are Demonstrated*, highlights the two most ambitious claims of his philosophy. Finally, in his last major work, *Passions of the Soul*, Descartes recapitulates some of the basic physiology from *Le Monde* and explores emotion and the means by which the body and soul interact.

'Razing Everything to the Ground':

What preceded was essentially a summary of orthodox Descartes, which the author now intends to expand, qualify, and revise in order to present a more accurate and univocal representation of Cartesian philosophy as it pertains to the mind-body duality. Desmond Clarke points out the danger of reading Descartes through the lens of a single work[9]; a concern which pays tribute to Descartes' own style of hyperbolic doubt. Clarke maintains that taking substance dualism, the notion that mind and body are necessarily separate because they have a different essence, as a first principle and retrofitting it to Descartes' earlier works where such metaphysical questions were not yet even posed obscures his intentions and elevates the mind-body dichotomy above its rightful place in the entire consortium. As a consequence of four centuries of distortion, Clarke identifies a situation where we have "the emergence of Descartes the metaphysician who was a substance dualist, rather than the Descartes the natural philosopher who flirted briefly with substance dualism only when dutifully making his contribution to the Catholic Counter-Reformation[9]." Such religious influence is readily evident; in response to Galileo's trial, Descartes refrained from publishing his own scientific treatise, *Le Monde,* and destroyed many of his other writings. Gordon Baker and Katherine Morris make a similar case in *Descartes' Dualism*[4], maintaining that the modern school of thought known as Cartesian Dualism would be unrecognizable to the man who gave it his namesake. They agree with Clarke that "Descartes' discussion of theological and metaphysical issues" is "the engagement of a reluctant participant with the politically dominant ideologies of his time[9]" rather than an integral part of his theory of mind.

Baker and Morris also highlight another potential pitfall, and like them, "[the author] refrain[s] from playing the popular party game of constructing refutations of Cartesian dualism.[4]" Rather than repeat the obvious superficial objections to a mind-brain duality – the problem of a non-physical entity causing motion, the question of the physiological locus for interaction, or the "ghost in the machine problem," or to look for logical fallacies in Descartes' reasoning, this investigation aspires to the more noble aim of showing that what served as an appropriate placeholder in the seventeenth century is no longer necessary in the twenty-first due to a radically different scientific and social landscape. This shift is akin to changing the rules of a game; it redefines the boundary between sense and nonsense, and determines what questions are permissible and relevant[4]. Thus to conduct a sensical analysis, one must maintain awareness of this evolution of thought and abandon those questions that no longer meet these criteria and reframe the essential ones so that they do. Similarly, the author does not intend to criticize the trivial points of Descartes' science and philosophy – that animals do not feel pain, the anatomical localization of the soul in the pineal gland, or the thermodynamic model of blood circulation. Such an endeavor would be futile and would distract from the primary object to establish a revisionist but true Cartesian theory of mind that can be juxtaposed with modern neuroscience without being wholly anachronistic.

Descartes himself was utterly aware of the possibility of being misrepresented. He writes in a letter to Chanut, "A certain Father Bourdin thought he had good reason to accuse me of being a sceptic, because I refuted the skeptics; and a certain minister tried to argue that I was an atheist, without giving any reason other than the fact that I tried to prove the existence of God[4]." Ignoring for the moment Descartes' particular religious claims and focusing solely on the irony, one can readily ascertain Descartes' expectation that his work would be received with great difficulty even in his own time, nevermind after a four century game of telephone. In the dedication to *Meditations* he recognizes that "although [he] believes [his] arguments to be certain and evident, still [he is] not thereby convinced that they are suited to everyone's grasp[17]." He is "fearful that many people will not be capable of adequately perceiving them… …because they demand a mind that is quite free from prejudices[17]." He explicitly identifies the mind-body duality as one of the most elusive parts of his philosophy: "It does not seem to me that the human

mind is capable of forming a very distinct conception of both the distinction between the soul and the body and their union; for to do this it is necessary to conceive them as a single thing and at the same time to conceive them as two things; and this is absurd[9]" and thus begins the arduous task of attempting to understand the theory whose own proponent labels its conclusion as such.

The Cartesian Texts:

The present work traces the development of Cartesian philosophy approximately chronologically, beginning with the epistemological question first posed in the *Regulae* and rearticulated in nearly every one of Descartes' major works. Descartes sought to find a "new and stable basis for all knowledge[17]," be it scientific, philosophical, or theological knowledge. The impetus for this was an epiphany he had in 1619 whereby he sought to replicate the certainty and precision he found in mathematics with regard to all other knowledge. His methodology was essentially to erect a new system upon which conclusions could be made with absolute certainty by first rejecting all preconceived notions and existing opinions. In *Meditation One* he begins by acknowledging that he has accumulated a battery of false opinions over his lifetime, and resolves to "apply [himself] earnestly and unreservedly to the general demolition of [these] opinions[17]." He progresses through the permutation where it becomes adequate not to prove that his opinions are necessarily false, but simply to cast doubt on them, and then that only the most fundamental opinions upon which the others rest require this treatment; "because undermining the foundations will cause whatever has been built upon them to crumble of its own accord[17]." He first describes this process in *Discourse* where he identifies three fundamental principles that he intends to reject, the senses, reasoning, and the certainty of wakefulness. He asserts that "nothing was exactly as our senses would have us imagine[16]" and "resolved to pretend that everything that had ever entered [his] mind was no more true than the illusions of [his] dreams[16]."

As in *Discourse* Descartes' first target in *Meditation One* is sensory perception. In direct contrast to the established method of sensory information constituting the only certainty and deductive reasoning following to establish conclusions from the observed world, Descartes maintains that the information from our senses cannot be assumed

reliable. Nor can one assume that he is awake; "there are no definitive signs by which to distinguish being awake from being asleep[17]," and if that is the case, then the objects of sensory perception do not correspond to objects in the physical world but rather the contents of dreams. He then questions whether the representation of objects in a dream necessitates their existence or at least potential for existence; i.e., is it possible to form a representation of something that does not otherwise exist? The next conclusion is that there is sufficient doubt for him to postulate that corporeal things do not exist at all; "body, shape, extension, and movement are all chimeras[17]," and he reaffirms the conclusion of *Discourse* even more adamantly at the end of *Meditation One,* "everything I see is false[17]."

Thus Descartes' search for an adequate first principle begins by casting doubt on one's most fundamental beliefs. Before turning to what he ultimately concludes can be accepted without doubt, his own existence, the core of the *cogito* argument, consider two features of Descartes' method of radical doubt. Though Descartes postulates that his sensory information is deceptive, he cannot be sure that he is awake, and his body nor any other corporeal objects exist, he does so as a logical exercise. This may seem obvious but one does not expect him to retain these conclusions, rather the purpose of rejecting these beliefs is to ultimately reinstate them on more stable foundations, and the conclusion of the *cogito* is not that the only certainty is that a non-corporeal self can be assumed to exist, but rather that this is the only stable first principle that can be taken as self-evident. The second important observation is the context in which Descartes begins his search. While the method is most completely articulated in the metaphysical context of *Discourse* and *Meditations*, it is a ubiquitous feature of Descartes' writing, and he declares at the beginning of *Meditations*, "I realized that once in my life I had to raze everything to the ground and begin again from the original foundations, if I wanted to establish anything firm and lasting *in the sciences*[17]." Though Descartes clashed with the empiricists on the reliability of sensory information, his initial foray into the search for knowledge was as a natural philosopher, and his method of hyperbolic doubt in this context can be seen as the addition of a preliminary step to the scientific method rather than a radically different process.

Descartes the Natural Philosopher:

At this juncture a survey of Descartes the scientist, or natural philosopher, proves advantageous. Descartes' scientific writing, very little of which is extant, and which is comprised of *Le Monde*, *A Treatise on Man*, *A Description of the Human Body*, and *The Dioptrics*, was chiefly concerned with physics, physiology, and animal behavior. Rather than to extensively catalog Descartes' largely supplanted scientific theories, this brief survey intends to highlight the context in which they emerged and the affinity between the particular scientific domains he investigated and the theory of mind that emerges out of the composite of his scientific and philosophical writing.

Cartesian mechanics* maintained that matter can only be set into motion by other moving matter and rejected the concept of action at a distance. Note simply here that Descartes did the majority of his scientific writing between 1620 and 1640; Isaac Newton was born in 1643. Moreover, Descartes' understanding of matter, which for him is the fundamental distinction between corporeal substances and mental substances, must be understood as pre-classical mechanics, pre-atomic theory, and thus cannot be held to today's much more rigorous definition. Descartes' work in physics and his formulation of a theory of matter also renders it unlikely for him to later maintain that a non-physical entity causes bodily motion; he explicitly denies a strict causal role for the soul in moving the body in *Passions*, but the juxtaposition of these two ideas serves the role of reinforcing the ultimate place for mind-body dualism in a coherent representation of Descartes.

Just as the Cartesian theory of matter preceded Newton mechanics, Descartes attempted to explain physiology before the invention of the light microscope. By the standards of his time, Descartes' physiology was actually quite impressive. He accurately described digestion and absorption, antagonistic muscular contraction and relaxation, and circulation of the blood. With respect to the latter, Descartes expanded William Harvey's explanation, and accurately described the path blood traverses from the right ventricle through the pulmonary and systemic circulation back to the right atria, though he attributed the motion to be caused by rarefication resulting from thermal function of the heart rather than mechanical pumping. Though by Descartes' time the

* As in the subdiscipline of physics

consensus was that the brain was the anatomical locus of mental function[†] and Aristotle's attribution of this to the heart was largely supplanted, Descartes contributed to denying any role for the heart other than the circulation of blood in *Le Monde* and *Passions*.

Descartes' crude physiology actually contributed precursors for many systems and mechanisms that were impressive first approximations considering the absence of microscopy or an understanding of cell theory, and many of his ideas, particularly regarding neural and endocrine function, have only been revised with regard to precise mechanism rather than underlying theory. Descartes understood the function of the circulation as a transport system between different parts of the body; the agents being transported he termed *animal spirits*. The naïve interpretation is to consider this consistent with Gilbert Ryle's classification of Cartesian dualism as a "ghost in the machine[9]," but a brief consideration of the etymology of the phrase reveals the contrary. 'Animal' refers to mechanical, or pure physiological function, i.e. capacities humans share entirely with other animals, and 'spirit' is to be understood as simply a fluid medium, in the sense that wine is a spirit, excluding any connotations to non-corporeality, and Descartes' animal spirits can be understood as proto-endocrine hormones rather than some abstract theoretical entity.

Descartes described nerve signaling as completely mechanical; he supposed that afferent and efferent neural transmission was the consequence of physical tension in nerve fibers that originated at either the brain or the periphery and effected a response at the other end, "in the same way in which, when we pull one end of a cord, we make the other move[19]." While this explanation is devoid of the neurites, action potentials, and neurotransmitters we understand today to be the mechanism of such signaling, these concepts could not have even been conceived without the basic understanding of the cell. A remarkable feature of Descartes' explanation is his speculation, which he makes with specific regard to pain transduction, that this organization of nerves allows perception of a stimulus by the brain interpreted to originate at the peripheral terminus of a nerve fiber, to in fact arise anywhere along the nerve and still be perceived as such. This property has

[†] The phrase mental function is used loosely here as a general term taken to include sensation, perception, control of motion, intellect, emotion, language, learning, memory, executive function, rational and abstract thought, autonomic and vegetative function; i.e., anything attributed to the brain, without consideration of the historical or theoretical context. A more precise definition is not critical at this point.

been readily demonstrated in conditions such as trigeminal neuralgia or phantom-limb syndrome.

Descartes also attempted to describe sensory and motor function. His account of sensory transduction, which will be explored in greater detail in the context of sensation as constituting one of the four modes of thought‡ in Descartes' theory of mind, contains nearly all of the features of the modern cell-based explanation whereby a physical stimulus is converted into a neural representation and relayed successively through a series of brain structures. Descartes identified seven senses, the five external senses and two internal senses, one of which constitutes sensations from the viscera, such as hunger and thirst, and the other the passions or emotions. Sensation corresponds to movement at the sensory organ being translated via Descartes' mechanical nerve to movement of a corresponding brain structure; motor function corresponds to the reciprocal descending pathway whereby movement of a brain structure results in the movement of muscles. In Descartes' system, there is a continuity between circulation and nerve conduction, and the animal spirits travel from the heart to the brain, where they are separated from the blood and then travel through the nerves, which are "little filaments or little tubes, which all come from the brain[19]."

It is also worthwhile to note that Descartes' investigation of *human* mental capacities, namely sensation, memory, and imagination, led him in 1632 to the dissection of *sheep's* brains and the subsequent study of animal behavior[9]. Descartes' work as a naturalist described in *Le Monde* was motivated by a desire to understand human behavior, and in a pre-Darwinian scientific landscape, his implicit recognition that non-humans could serve as illustrative model systems, or at least some his acknowledgement of some degree of mechanistic homology, suggests that he viewed non-human and human mental function along a continuum rather than fundamentally different.

Some of the finer parts of Descartes' neurophysiology begin to break down when concepts such as cell theory, synaptic transmission, and the blood-brain barrier are anachronistically layered on top of them, but again the point is not to subject Descartes' science to critique by today's standards, but rather to note four hallmark features of his role as a natural philosopher. First, there is a remarkable harmony between his

‡ Sensation, imagination, memory, and pure intellect; discussed below in 'A Thing that Thinks'

descriptions of many physiological phenomena given the constraints that he was operating under and our present explanations for the same phenomena. To progress from animal spirits to electrochemical signaling to explain nerve transmission is not the same type of jump as to explain replace humoral imbalance with external pathogens to explain disease. Second, Descartes' affinity for neuroscience, or whatever one would call its seventeenth century precursor, must remain an essential ingredient in constructing a composite Cartesian ideology; the increasing emphasis on metaphysics and theology in his later works cannot simply supplant this observation but must be considered in concordance with it. Third note Descartes' goal remains to *explain* as many phenomena as he can. He broke with the Aristotelian tradition of simply attributing capacities to substantial forms; Moliere satirizes this in *La Malade imaginaire*, where Doctor Bachelierus 'explains' that opium makes one fall asleep because of its "dormitive power.?" For Descartes, such non-explanatory restatement is grossly inadequate. For now it suffices to simply highlight his explanatory goal. Finally, the fourth critical feature of examining Descartes as a scientist is to highlight the dynamic nature and constant revision of scientific theory. Consider Descartes' mechanics as an example: motion of matter is caused by the motion of other matter. A corollary of this requires that to explain action at a distance, which is readily observable, particulate matter must occupy all space and this precludes the existence of a vacuum[§]. Let us briefly trace the evolution of this idea over the next four hundred years from Newtonian mechanics to relativity to quantum mechanics; the explanations become increasingly more precise and general as the scientific background changes. Classical Newtonian mechanics is a fine first approximation as long as the object is not moving near the speed of light; until the introduction of relativity, such a phenomenon could not be otherwise explained, but Newtonian mechanics was an adequate temporary placeholder until the theory could be sufficiently refined to accommodate new data. Descartes was well aware of this quality of scientific explanation, and actually explicitly acknowledges this in *Le Monde*. He distinguishes whereby eventually a theory reaches a point where it must temporarily cede to theoretical constructs that it cannot in and of itself yet explain. Noting this feature in regard to Cartesian science allows one to pose two questions. First, is it possible to

[§] Descartes actually explicitly states this in *Le Monde*

extrapolate this to Descartes' theory of mind and suppose that mind-body dualism is a placeholder for aspects of human mental capacity that Descartes recognized he could not sufficiently explain otherwise? Second, is there an upper limit to what can ultimately be explained mechanistically, a question that could not be seriously framed in Descartes' time, but since the conceptual landscape has been redrawn and the rules of the game have changed, one is now in a position to consider?

The *Cogito*:

Continuing along a chronological journey through Descartes' writings, having ascertained a means to establish 'firm and lasting' knowledge by beginning with hyperbolic doubt, rejecting all preconceived notions, in order to recreate the underlying order and certainty of mathematics in other domains, and returning from a brief digression through Cartesian science, we now search for a satisfactory first principle upon which to build everything else. This leads to the crux of Descartes' metaphysics, and perhaps of the whole unified ideology we attempt to create – the *cogito* argument, first presented in *Discourse*, discussed at length but not nominally in *Meditations*, and then recapitulated in *Principles*.

In *Discourse*, after Descartes has rejected his sensory perception, reasoning, and his wakefulness, he arrives at the one principle he cannot cast sufficient doubt on and cannot reject, making the first statement of the *cogito* argument,

> "But immediately afterward I noticed that, during the time I wanted thus to think that everything was false, it was necessary that I, who thought thus, be something. And noticing that this truth – *I think ,therefore I am*–was so firm and so certain that the most extravagant suppositions of the skeptics were unable to shake it, I judged that I could accept it without scruple as the first principle of the philosophy I was seeking[16]."

That he was capable of, or actively engaged in, questioning the validity of these principles – sensory perception, reasoning, wakefulness, necessitates that some entity, i.e. himself, is performing this thought, and in order to do so must exist, irrespective of whether sensory information is reliable, subsequent reasoning is efficacious, or he is awake or dreaming. This argument necessarily arises out of Descartes' method of

hyperbolic doubt; the very act of doubting, which we can consider a subtype of thinking, forces one to acknowledge that one is doing so. Try the mental exercise of trying to cast doubt on the fact that you are in fact doing so and you will readily discover "it is not possible for us to doubt that, while we are doubting, we exist; and that this is the first thing which we know by philosophizing in the correct order[18]." Moreover this is the only conclusion that passes the rigor of Descartes' evaluation,

> "while rejecting in this way all those things which we can somehow doubt, and even imagining them to be false, we can indeed easily suppose that there is no God, no heaven, no material bodies; and even that we ourselves have no hands, or feet, in short, no body[18]."

He continues in *Principles*, "yet we do not on that account suppose that we, who are thinking such things, are nothing: for it is contradictory for us to believe that that which thinks, at the very time when it is thinking, does not exist[18]." The conclusion follows, "accordingly, this knowledge, *I think, therefore I am*, is the first and most certain to be acquired by and present itself to anyone who is philosophizing in the correct order[18]."

In both appearances of the *cogito*, note the appearance of temporally qualifying phrases that link the argument to the present. In *Discourse*, 'during the time I wanted thus to think that everything was false,' and in *Principles*, 'that which thinks, at the very time when it is thinking.' The *cogito* does not say I think, therefore I *was* or *will be*, but I *am*, and extrapolation beyond the instant in which one is thinking, a precondition for concluding the immortality of the soul, is not inherent in this argument.

The other characteristic feature of the development of this argument is its utter logical simplicity and apparent necessity, as the mental exercise of trying to recreate it reveals. Descartes attempts to entertain the contrapositive, "on the other hand, had I simply stopped thinking… …I would have no reason to believe that I existed[10];" the conclusion is the same and equally self-evident. While he accepts the *cogito* as a first principle because it is the only thing remaining after hyperbolic doubt, "doubtless I did exist, if I persuaded myself of something[10]," Descartes uses a different feature of the argument to subsequently extrapolate it. He notices that "this truth – *I think, therefore I am*–was so firm and so certain that the most extravagant suppositions of the skeptics were unable to shake it[18]," and therefore concludes that "things we conceive very clearly and

very distinctly are all true[18]." One could dwell on the apparent non-sequitor here, but for us it suffices to simply note that Descartes supposes truth from clarity and distinctness.

The final task in explicating the *cogito* proper is to qualify the 'I' that Descartes' refers to at this stage in the argument. "We ourselves have no hands, or feet, in short, no body, yet we do not on that account suppose that we, who are thinking such things, are nothing[18]." The very conditions that allow the *cogito* to emerge prevent one from positively affirming the existence of a corporeal body. Descartes breaks the logical progression down in *Meditations*. First, "thought exists, it alone cannot be separated from me;" and it follows that "I am; I exist – this is certain." Unifying these two, "I am therefore precisely nothing but a thinking thing[17]," which, "in order to exist, needed no place and depended on no material thing[16]." From this perspective, the mind-body duality that emerges out of the *cogito*, is a logical consequence, and this author proposes to some extent an artifact, resulting from a necessary concession that the *cogito* makes in order to achieve its absolute certainty. Descartes never intends to permanently cast doubt on the reality of corporeal bodies, and in fact he acknowledges the "commingling of the mind and body[17]" even at this early stage in the development of the argument. Similarly, this author maintains that the mind-body duality and accusation of substance dualism is overemphasized because Descartes fails to reconstitute the self as an essential entity, or *ens per se*, consisting of the essential union of the *res cogitans* established by the *cogito* and the *res extensa*, or body, with the same certitude with which he distinguishes them, a difficulty he was undoubtedly aware of – "it is necessary to conceive [mind and body] as a single thing and at the same time to conceive them as two things; and this is absurd[9]."

Mind-Body Dualism:

Regardless of the extent to which Descartes ultimately advocates this view, the immediate consequence of the *cogito* is a fundamental distinction between the *mind*, or alternatively the thinking self whose existence is affirmed by the cogito, or soul, and the *body*, or extended self. We entertain this distinction proper as it emerges from the Cartesian texts before progressing to the residual question that restricts the domain of the latter and deals solely with a mind-brain duality. Essentially the direct implication from the *cogito* is that what can be absolutely affirmed to exist is a 'thinking thing,' because its

existence is guaranteed by the undeniable observation that it is in fact thinking. This sounds remarkably like the scholastic notion of non-explanatory substantial forms, which we have already depicted Descartes as explicitly rejecting. The traditional argument takes from the *cogito* that "I am therefore precisely nothing but a thinking thing[17]," which, "in order to exist, needed no place and depended on no material thing[16]," and identifies the mind as a substance defined[**] by thought, and in contrast to the body,

> "all that is capable of being bounded by some shape, of being enclosed in a place, and of filling up a space in such a way as to exclude any other body from it; of being perceived by touch, sight, hearing, taste, or smell; of being moved in several ways, not of course, by itself, but by whatever else impinges on it[17],"

defined by extension. Descartes repeats this delineation in *Principles*, "extension, or figure, or local motion (or any similar thing which must be attributed to a body) does not belong to our nature[18]," where at this juncture 'our' refers to the mind or 'thinking self' of the cogito. Even Clarke, the most dogged critic of attributing substance dualism to Descartes concedes that "some mental properties are not reducible to the properties of Cartesian matter[9]." While a preliminary distinction is real, its overemphasis gives way to the two-worlds view that Baker and Morris criticize[4], relegating the mind and body to separate independent spheres because of their different natures. A preview of the implication of this two-worlds view is that mind and body necessarily operate wholly independently, and an attempt to reconcile this with Descartes' subsequent insistence of their interaction forces one to choose between causal dualism, whereby "mental states and states of the body are logically independent but causally interrelated: causal interaction is, as it were, the glue bonding mind to body in each individual person[4]" or occasionalism, in which "these two streams of events are kept marching in step by God's constant interventions[4]." The former violates Cartesian physics by requiring that a non-corporeal entity, the mind, causes the motion of matter, i.e. the body, whereas the latter

[**] 'Defined' is intended to signify that the property that follows (here thought, subsequently for body, extension), refers to the essential nature of the substance being discussed in a Scholastic fashion; i.e., thought is the single defining property which makes the mind a mind. Too much emphasis is not made to define all of the Scholastic concepts involved – substance, mode, form, because a detailed discussion of this is peripheral to the present discussion, but a more detailed account can be found in Chapter 8 of Clarke, Chapter 5 of Baker & Morris, and throughout Sepper.

merely displaces the burden of explanation. To this author neither is satisfactory, and therefore the premise that prompted the either/or dilemma is resubmitted to examination.

Aside from these two different natures or essences that render mind and body fundamentally different, the other distinction that emerges directly from the *cogito* is the different quality of one's understanding of their existence. In *Discourse* Descartes claims "I could pretend that I had no body and that there was no world nor any place where I was, but that I could not pretend, on that account, that I did not exist[16]." He continues, "this "I," that is, the soul through which I am what I am, is entirely distinct from the body, and is even easier to know than the body, and even if there were no body, the soul would not cease to be all that it is[16]." Three features prove striking about this second claim. It is the first point at which Descartes equates the *res cogitans* with the soul. His claim that the soul is 'even easier to know than the body' echoes his conviction that "things we conceive very clearly and very distinctly are all true[18]." Finally the last part of this passage does not necessitate the immortality of the soul, but only its independence from the body.

Before considering the nature of their interaction, one must attempt to understand the logical relationship between the self/soul/mind/*res cogitans*, the corporeal physical body/*res extensa*, and their composition, if the latter can be said to exist. Isolation of the mind as an independent entity has already been satisfactorily addressed. Descartes makes the same explicit claim for the body.

> "I also acknowledge that there are other faculties, such as those of moving from one place to another, of taking on various shapes, and so on, that, like sensing or imagining, cannot be understood apart from some substance in which they inhere, and hence without which they cannot exist[17]"

We omit tracing his proof of the existence of corporeal bodies, and instead return to Descartes' earlier challenge to form "a very distinct conception of both the distinction between the soul and the body and their union[9]." In *What Am I?* Joseph Almog attempts to logically prove this simultaneous distinction and union[3]. He defines DM, Descartes' mind; DB, Descartes' body; and RD, Rene Descartes, the composite entity. The proof of distinction operates on conceivability and numerical distinction; DB has property x; DM does not, therefore DB ≠ DM. The proof of independence is even more vacuous, and

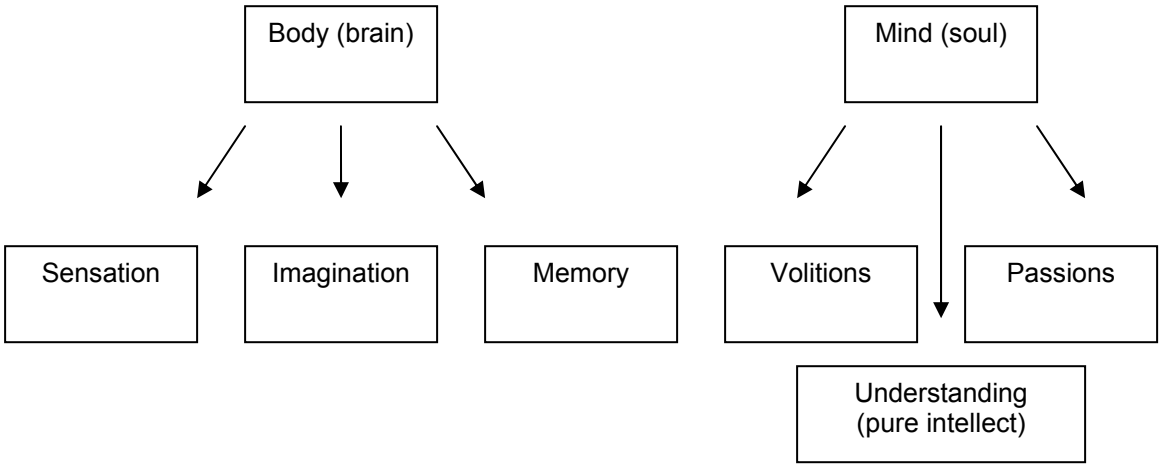
though logically sound, its practicability is not particularly satisfying to this author, who retains Descartes' view that "to conceive them as a single thing and at the same time to conceive them as two things is absurd[9]." The usefulness of Almog's discussion is that he elegantly articulates the logical relationship between these entities, that RD consists of the union of DM and DB, that DB and DM are independent of each other, and both are contingent but not wholly dependent on RD.

Another fundamental difference used to demarcate the realms of mind and body is the question of their divisibility. A body, being composed of particulate matter, is inherently divisible. Descartes maintains "on the other hand, the mind is utterly indivisible[17]." The faculties of the mind, "willing, sensing, understanding, and so on" cannot "be called 'parts' of the mind, since it is one and the same mind that wills, senses, and understands[17]." For Descartes "this consideration alone would suffice to teach [him] that the mind is wholly diverse from the body[17]." Should we succeed in attributing these faculties of the mind, willing, sensing, and understanding, to the physical brain, and identifying a spatial locus for each of them via lesion studies or neuroimaging, this claim would evaporate.

The final means for delineating between mind and body is a division of labor of sorts. In the above passage Descartes assigns to the mind the functions of willing, sensing, and understanding. The faculty of sensing remains to be qualified, as this is a faculty shared by the Cartesian mind and body, albeit in different capacities. However Descartes has quite precisely assigned nearly every conceivable mental[††] capacity to either the mind or the physical brain. This is best represented

[††] We now transition into specifically considering the brain and its distinction from the mind rather than the whole body. Obviously physical functions concerning the entire body proper are not conceivable to consider attributing to the mind; the mind does not walk (though we do include control of locomotion)

graphically:



This representation attempts to integrate the various descriptions of Descartes' assignment of different mental faculties to either the soul or the body. In *Passions*, he claims "there is no better path for arriving at an understanding of our Passions than to examine the difference between the soul and the body, in order to understand to which of the two each of the functions within us should be attributed[19]." The organization of his division scheme is sensation, imagination, and memory are variations of the same essential process which differ only in their input. These are faculties shared with non-human animals and therefore necessarily assigned to the body. Understanding, or pure intellect, represents an increasing degree of abstraction of any of these three processes, which Descartes assigns to the soul. Together these four comprise 'thought' or *cogitatio*. The two functions he ascribes wholly to the soul are volitions, or actions of the will, and passions, which essentially are emotions, though in *Passions* he allows that the soul is informed by the body with regard to these passions, and they may have their terminus in either the soul or the body. Despite these ambiguities and some inconsistency between different texts or different passages within the same text, Descartes does attempt to circumscribe the functions of the soul and the body.

'A Thing that Thinks':

The *cogito* establishes Descartes, or any man for that matter, as a thinking thing. He questions, "but what then am I? A thing that thinks. What is that? A thing that doubts, understands, affirms, denies, wills, refuses, and also that imagines and senses[4]," attempting to provide a comprehensive definition for *cogitatio*. Let us examine this definition in order to identify what exactly constitutes Cartesian thought. The first, doubt, is necessarily included because it was the means by which we arrived at this conclusion in the first place. Understanding we will loosely equate with awareness, conscious perception, and the highest degree of abstraction of other modes of thinking, i.e. the pure intellect that Descartes ascribes to the soul. Affirmation and denial and willing and refusal represent pairs of opposites that relate to the volitions, again a function of the soul. Imagination and sensation however, are explicitly assigned to the body[17], and here are undeniably included in a definition of thought. ["Sensing,] is nothing other than thinking" and the "power of imagining depends on something distinct from me[17]," i.e. a body. Contrast this account from *Meditations* with that in *Principles,* where "there are only two modes of thinking in us; that is, the perception of the intellect and the operation of the will[18]." Cartesian thought can then be taken to include sensation, imagination, memory, pure intellect, and operation of the will, where the first three of these are faculties of the body, the fourth comprises perception of any of the first three by the soul, and the fifth is an exclusive function of the soul.

We now consider each of these in turn. Sensation will be elaborated on at length[‡‡], but is included here for completeness and clarity. Essentially the Cartesian account of sensation corresponds to detection of physical stimuli, and interpretation and integration by the brain of features that really exist in the external world. This also includes the two internal senses; the commonality is the correspondence to corporeal reality and transduction via a physical sense organ. Baker and Morris make the useful distinction between bodily sensation, which they denote with a one subscript, $seeing_1$, and what they term rational, restricted sensation, which Descartes assigns to the soul, alternatively the conscious perception or judgment of sensing something, denoted by a two subscript, $seeing_2$[4].

[‡‡] In Sensation and Sensory Phenomena below

> "The same cognitive power [involved in sensation] can also be applied to images stored in the brain; when performing this function, it is called 'memory.' Finally , if it becomes creative and causes the occurrence of new brain patterns, it is called 'imagination' when it depends on previously stored images, and 'pure intellect' when it acts independently of perceptions or impressions[9]."

Clarke's paraphrase is particularly elegant because it distinguishes each capacity according to its object, but the original is essentially the same, "the same power therefore is called 'pure intellect,' 'imagination,' 'memory' or 'sensation,' depending on its different operations[9]." Imagination, thus, is analogous to sensation except that the objects that the brain operates on are not readily available to the sense organs and some amount of creativity is permissible. One can imagine, for instance, an object previously perceived by the senses but not at the moment present in the environment; imagine an apple. Similarly, the imagination is used in constructing composite realities that escape the necessary temporal constraints of the senses, a song from notes separated in time. One can imagine a griffin by fusing the separate images of a lion and an eagle. The imagination also allows for a certain degree of abstraction, but is restricted by its ultimate dependence on prior sensory exposure. For example, one can readily imagine a triangle, but cannot imagine a chiliagon, a thousand-sided figure[9]. While such a figure can certainly exist and is conceivable, one cannot form an image of it that can be easily distinguished from say a nine-hundred-ninety-nine sided figure. Clarke maintains that the imagination limits intellectual abstraction to what is conceivable, though in this case the constraint is simply the capacity to form an image[9]. Sepper notes that in Descartes' earlier works, the imagination, or *phantasia*, is the theoretical precursor to what eventually becomes the intellect[39]. Even here, the only difference is a gradation of the degree of permitted abstraction, hardly the distinction that demands fundamentally different domains. This shared mechanism and seemingly fluid boundaries between separate mental capacities that are on the one hand rigidly assigned to either body or mind should ultimately provide us with enough confidence to localize them all entirely to the brain and eliminate the need to hypothesize a non-corporeal mind.

Descartes deals with memory in much the same way. Whereas imagination can use prior sensory exposure in a creative fashion, memory operates on 'images stored in the brain.' Consider Descartes' explanation of memory,

> "imagine that after issuing from gland H spirits pass through tubes 2, 4, 6 and the like, into the pores or gaps lying between the tiny fibres making up part B of the brain. And suppose that the spirits are strong enough to enlarge these gaps a little, and to bend and arrange any fibres they encounter in various ways, depending on the different ways in which the spirits are moving and the different openings of the tubes into which they pass. And they do this in such a way that they also trace figures in these gaps, corresponding to those of the objects. At first they do this less easily and perfectly here than on gland H, but they gradually improve as their action becomes stronger and lasts longer, or is repeated more often. Which is why in such cases these patterns are no longer easily erased, but are preserved in such a way that the ideas that were previously on this gland can be formed again long afterwards without requiring the presence of the objects to which they correspond. And this is what memory consists in[39]."

Descartes explanation identifies memory as the disposition of animal spirits to repeat a particular trace that they have already performed due to changes in the network resulting from continued activation of this trace. While physical deformation of the fibers has since been replaced by NMDA receptors, Descartes identification of a particular brain trace rather than a locus precludes the need to hypothesize an engram, and is remarkably consistent with our present understanding of memory. Similarly, in *Passions*, "these traces are nothing but this: the pores of the brain through which the spirits have previously made their way because of the presence of this object[19]." Like sensation and imagination, memory is explained mechanistically entirely in terms of the physiological brain. Descartes does make room for a capacity called intellectual memory, apparently a function of the soul, but Clarke makes a convincing case that this appears as an afterthought appended to a discussion of body-based memory, and it is only acknowledged but not explained in response to the question of whether in the afterlife one is capable of remembering experiences during his or her lifetime[9].

The final mode of this cognitive power is pure intellect, 'which acts independently of perceptions or impressions.' Pure intellect or pure understanding is conscious thought independent of a corresponding stimulus or image. Consider the example of the triangle versus the chiliagon; the former uses the imagination, the latter

pure intellect. Clarke argues that 'pure' refers to thought that is completely within ones own control as a thinking being and not constrained by the realities of sensation, imagination, or memory[9]. He maintains that this independence characterizes pure intellect in contrast to other forms of thought. Moreover he adds that this only *requires* separation from the physiological brain when substance dualism is read anachronistically into the argument; Descartes' discussion of pure intellect preceded the *cogito* by nearly twenty years. While he does assign pure intellect to the soul, the continuity between this capacity and the others should cast further doubt on a fundamental distinction that requires this.

Sensation:

Sensation is perhaps the most interesting form of thought that Descartes describes. On the one hand, "[sensing,] precisely so taken is nothing other than thinking[17]" and on the other it "surely does not take place without a body[17]." Using vision as an example, we consider again Baker and Morris' distinction between $seeing_1$ and $seeing_2$. The former refers to the physical, anatomically-dependent act of transducing a visual stimulus from the environment to a brain pattern. The latter describes the conscious perception of seeing, which Baker and Morris maintain that according to Descartes' account, precisely involves making a judgment, an action of the soul, that one is $seeing_1$ something. Thus the statement 'I see_2 light' is logically equivalent to 'I judge that I see_1 light.' A parallel feature of the *cogito* emerges here; while one can be mistaken in the statement 'I see_1 light,' the sincere assertion that 'I see_2 light' is necessarily true. Moreover, $seeing_1$ does not necessitate $seeing_2$, as in blindsight[§§], and $seeing_2$ does not necessitate $seeing_1$, as in visual hallucinations.

The contrasting accounts create a complex picture. Descartes insists that "I could not sense any object unless it was present to a sense organ. Nor could I fail to sense it when it was present[17]," grounding the senses in the physical world and localizing them to the anatomical brain, but he by no means here insists on their infallibility. In fact, he maintains that our senses readily deceive us, such as in perceiving towers in the distance as round when they are in fact square. This applies equally to the internal senses; he

[§§] Blindsight is discussed at length below.

discusses phantom limb syndrome and notes that someone with dropsy will feel thirsty when in fact water will be harmful to them. This final example is less robust, because the physical sensation of thirst is no different, merely the question of its contribution to welfare, which Descartes identifies as the purpose of our senses. The other side of this account is a return to the primacy of sensation$_2$; "what I thought I had seen with my eyes, I actually grasped solely with the faculty of judgment, which is in my mind[17]." The distinction Descartes makes between sensing$_1$ and sensing$_2$ is important and certainly still relevant, and is perhaps equivalent to the modern distinction between sensation and (conscious) perception. What the author questions is whether there is a discrete boundary that necessitates fundamental separation or whether like the other faculties surveyed thus far, these phenomena range along a continuum and four hundred years of brain science have allowed us license to revise Descartes' division and eliminate the need to hypothesize a non-physical entity responsible for the highest levels of processing. Descartes offers as an example our perception of a star versus a torch[17], and notes that "there is no real or positive tendency in my eye toward believing that the star is no larger than the flame[17]." For Descartes, this *judgment*, is 'perceived through the mind alone,' though we have now located mechanisms for interpreting size in relation to distance early in the visual pathway.

Descartes suggests that the highest levels of processing are faculties of the soul. Using vision as a representative example, we consider his tentative pathway beginning with transduction to attempt to ascertain precisely where the soul takes over from the physiological brain. He understands the general principle of transduction, likening the action of physical stimuli on the sense organs to "the same way as wax receives an impression from a seal[9]." Descartes likely did not realize this, but the analogy of temporary reversible physical distortion corresponds directly to bleaching of photoreceptors, vibration of the basilar membrane, activation of taste and olfactory receptors, or bending of mechanoreceptors. With regard to the visual system, "objects of vision are communicated to us by the mediation of transparent bodies between them and us[19];" Descartes is effectively talking about photons here. These 'transparent bodies' then "locally move the little filaments of the optic nerves at the back of our eyes[19]." The 'little filaments' that he discusses have already been identified as nerves, and thus

Descartes observes that a physical stimulus is converted into a neural signal. More generally, "[all of the senses] excite some movement in our nerves, which passes to the brain by means of them[19]."

The sensory nerves then move "the parts of the brain these nerves come from[19];" i.e. there are structures in the brain that receive the sensory input. Continuing Descartes' account from *Passions*, "they move them in as many different ways as there are diversities they make us see in things, and that it is not the movements occurring in the eye, but those occurring in the brain, that immediately represent those objects to the soul[19]." From this one can glean that he understands a sequential relay and physiologically based cognitive processing of sensory information. An earlier account from Rule 12 in *Regulae* follows a slightly different path; impressions received by the sense organs are conducted to the *sensus communis* or 'common sense' then to the *phantasia* or imagination, and then finally to the *vis cognoscens* or "the force through which we properly know things[20]," only the latter of which is not physically grounded. Merging these two accounts, Descartes' 'common sense' can be surmised as a precursor to the thalamus; in his pathway it integrates sensory information but precedes conscious perception. As presented here, Descartes' account does not seem to necessitate that one attribute conscious perception entirely to the non-corporeal soul; he explicitly affirms a physiological locus for the *phantasia*.

While Descartes' sensory pathways do ultimately converge on the soul; "these various movements of the brain make our soul have various sensations[19];" he also notes a purely bodily pathway whereby "they can also, apart from the soul, make the spirits take their course toward certain muscles rather than others, and so make them move our members[19]." This allows for several interpretations ranging from a simple reflex to something analogous to the retinotectal projection (whereby visual information bypasses primary visual cortex and we are not explicitly aware of seeing something but can still respond to it) to the supposition that consciousness does not require the extra-body soul. The author prefers the third but does not insist that readers subscribe to a particular interpretation of this feature.

The final critical feature of Descartes account of sensation is that he recognizes the brain's encoding of sensory information. He describes a one-to-one correspondence

but non-resemblance between external stimuli and brain patterns[9]. Descartes describes this phenomena in *Regulae* and *Le Monde*, drawing an analogy between words and the objects that they represent. Non-resemblance is critical here; in his work on the visual system, *The Dioptrics*, Descartes maintains that the image on the retina is not identical to that in the external world, "as if there were yet other eyes in our brain which we could perceive it[9]." Descartes' early recognition and emphasis on this feature of sensory perception carries several implications. First of all it underscores the extent of his scientific understanding. Moreover, this preempts and diffuses later attacks on the Cartesian account such as Ryle's 'ghost in the machine' and Anthony Kenny's accusation of a homunculus fallacy, whereby Descartes' apparent dualism is analogous to Aristotle's non-explanatory restatement, the most obvious implication of which is that it implies an infinite regress.

Functions of the Soul:

While on the one hand Descartes undoubtedly makes claims to a necessary separation of the mind and body and some form of dualism, there is a parallel current of increasing physiological explanation and ambiguity between the functions of soul and the brain. As one hones in on this boundary from different perspectives, the theme that should emerge is that Descartes attributes to the soul the most abstract and complex degrees of particular functions, thus far all of which in their simpler more concrete forms, are explained physiologically. Having illustrated this, we now investigate what, according to Descartes, belongs *exclusively* to the domain of the soul or mind.

There are three such functions that Descartes seems to attribute solely to the mind. In Article 17 of *Passions*, he states that

> "having thus taken into consideration all the functions that belong to the body alone, it is easy to understand that there remains nothing in us that we should attribute to our soul but our thoughts, which are principally of two genera – the first, namely, are the actions of the soul; the others are its passions[19]."

The so-called 'actions of the soul' are alternatively referred to as volitions, "come directly from our soul and seem to depend only on it[19];" this is the will. The 'passions of the soul' are "all the sorts of cases of perception or knowledge to be found in us[19]." In

Article 27 Descartes explicitly defines 'passions of the soul' as "perceptions or sensations or excitations of the soul which are referred to it in particular and which are caused, maintained, and strengthened by some movement of the spirits[19]." From this general definition, Descartes then variously classifies and organizes the passions, and eventually identifies a subclass of which there are six fundamental passions, wonder, love, hatred, desire, joy, and sadness, which he discusses at length, and the latter part of *Passions* reads like a theory of emotion or temperament. Thus the author divides Descartes' notion of passions into two distinct concepts, self-awareness or *conscientia* in general and the specific passions that Descartes enumerates and their products which appear to correspond to emotional responses. These, in addition to the volitions, comprise the entirety of the functions which can be attributed exclusively to the soul.

Self awareness, or *conscientia* is exemplified by the previous example of sensation with the two subscript. One can expand this beyond sensation and identify *conscientia* as the awareness of, or judgment that one is, performing a particular faculty of the body. To $walk_2$ makes as much sense as to see_2 and is not necessarily equivalent to $walking_1$. Essentially this faculty can be described as self-knowledge or reflective knowledge. Baker and Morris note this as a restricted case of infallibility, i.e. that "a person cannot be mistaken in his own (sincere) judgments that he has particular 'perceptions' or that he makes/has made particular judgments[4]." The judgment itself can be misinformed, but the second order judgment, or the judgment that one is judging thus, is necessarily correct. This function of the soul seems to be wholly included under our present understanding of consciousness, by no means a trivial component of human mental capacity, but one that is identified entirely with the physiological brain, namely, the cerebral cortex.

Descartes' more limited discussion of 'the passions of the soul' identifies their principle effect as to "incite and dispose their soul to will the things for which they prepare their body[19]," or alternatively to "dispose the soul to will the things nature tells us are useful and to persist in this volition[19]." Descartes identifies a subclass of passions caused by the soul "which decide to conceive of this or that object[19];" this seems to correspond to the *conscientia* described above. The remainder of the passions however, are caused by "the temperament of the body alone or by impressions haphazardly

encountered in the brain… …excited as well by objects which move the senses[19]." Thus the passions originate in the body, convey a particular sense to the soul, which effects or modulates a particular volition, which is then reflected back to the body resulting in an action; in Clarke, "their natural function is to stimulate the soul to consent and contribute to actions that can help to protect the body or make it more perfect in some way[9]." Since both the origin and terminus is in the body, this prompts the question of the particular role of the soul in this system and whether it can be regarded as a superfluous shunt. There is an implicit suggestion of this in Descartes; "These movements that are produced in the blood by the objects of the passions follow so promptly solely from the impressions made in the brain and the disposition of the organs, even though the soul contributes nothing at all to them[19]." This author is in agreement with Clarke's conclusion that this suggests that "this naturally instituted coordination between our sensations, the passions they generate, and the relevant behavioural response can operate without any intervention of the soul or mind[9]." Rather, here the soul is superimposed on an otherwise autonomous system whereby sensations exact actions as a black box for which Descartes deems the science of his time unable to satisfactorily account for.

The final function that Descartes ascribes to the soul is this second genera of our thoughts, which he alternately calls 'the actions of the soul, 'our volitions,' and 'the will,' which in contrast to the passions, "come directly from our soul and seem to depend only on it[19]." In the next article he distinguishes between volitions that terminate in the soul, the example he offers is to will to love God, and those that terminate in the body, for instance, to have the volition to walk. Clarke defines Descartes' will as the "distinctive power or ability that human agents have and in virtue of which some of their actions are subject to moral evaluation[9]." This carries two features which prevent us from simply identifying the will as the capacity to initiate voluntary action and relegating it to the prefrontal cortex and basal ganglia, thus effectively entirely dismantling Descartes' soul. First, Descartes maintains that the will is uniquely human, and second, this is the first appearance in our discussion of consideration of morality. However the motif continues here; consider this account of the will from *The Description of the Human Body*,

> "the soul can cause no movement in the body unless all the corporeal organs required for that movement are properly disposed. Besides, when the body has all

> the organs disposed for this movement, it does not need the soul to produce it. Consequently all those movements that we do not experience as depending on our thought must not be attributed to the soul but only to the disposition of our organs; and even those movements that are called 'voluntary' proceed principally from this disposition of the organs, for they cannot have been produced without it, no matter how much we will it[20]."

Descartes seems to assert here that the 'disposition of the organs' is necessary and sufficient for at least those volitions that have their terminus in the body. Two caveats remain – this says nothing about the class of volitions that terminate in the soul, and also, this passage is immediately followed by the ambiguous phrase "even though it is the soul that determines them[20]."

Thus Descartes presents three functions of the soul, self-knowledge or consciousness of one's thoughts and actions; translation of sensory or other bodily-derived stimuli into volitions; and finally volitions or willing to perform particular thoughts or actions. For each of these, we have painted a picture where in Descartes' grand scheme the function belongs exclusively to the soul, but as each is considered in detail, contradictory or at least ambiguous accounts of bodily autonomy and dependence on the soul are both asserted and the boundary becomes increasingly abstruse. Like picking up a jellyfish, the brilliance and clarity vanishes when one removes it from the water. The author does not attribute this feature to Descartes being deliberately esoteric or proposing an ill-informed theory, but rather maintains that his theory of mind was a work in progress up until his death and that much of the apparent ambiguity that we have identified corresponds to placeholders as he sought increasingly better explanations.

'The Little Gland':

That we have identified individual processes that jointly involve the body and the soul appears to require an interaction of some sort between these two entities. This claim is made here as general as possible; the term 'interaction' is not intended to presuppose any of the rigidity variously associated with the philosophy of mind concepts of 'interactionism' or 'causal dualism,' but simply to denote the observation that the actions and passions of the soul are conveyed to and from the body. Descartes readily concedes this; identifying the "power of the soul to move the body and the power of the body to act

on the soul[4]." This forces us to confront three questions concerning the quality, locus, and mechanism of this interaction. The soul by definition lacks extension, whereas the body is precisely defined by it, and the interaction between these two presents a philosophical problem that the present investigation does not intend to delve further into but merely points out for completion.

While "the soul is truly joined to the whole body, and that one cannot properly say that it is in any one of its parts to the exclusion of the others[19]," Descartes readily assigns a physiological locus for the interaction between body and soul,

> "the part of the body in which the soul immediately exercises its functions is in no way the heart; it is not the whole brain either, but only the innermost of its parts – a certain extremely small gland, situated in the middle of its substances, and so suspended above the duct by which the spirits of its anterior cavities are in communication with those of the posterior that its slightest movements can greatly alter the course of these spirits, and conversely the slightest changes taking place in the course of the spirits can greatly alter the movements of this gland[19]."

In *Passions* Descartes explains his reasoning for choosing the pineal gland as the 'principal seat of the soul.' He identifies it as the only unpaired structure in the brain, in contrast with the hemispheres, the thalamus, and the sensory organs, and concludes that "inasmuch as we only have a single and simple thought of a given thing at a given time, there must necessarily be some place where the two images… …can coalesce into one before they reach the soul[19]." Moreover Descartes incorrectly notes that this gland is unique to humans. Interestingly, modern pineal physiology may have caused Descartes to more readily affirm its role as the seat of the soul than to discount it. The pineal gland has now been identified as primarily responsible for producing melatonin, a hormone involved in regulating Circadian rhythms in response to photoperiod. Perhaps more interesting, the pineal gland shares developmental origin with the retina and is also not isolated from the circulation via the blood-brain barrier like much of the brain, rendering it capable of interacting with Descartes' animal spirits.

In an extension of Descartes' mechanical model of nerve transmission, communication between the pineal gland and the rest of the body is achieved by movement. Movements of the pineal gland direct the animal spirits in a particular way,

and the "animal spirits enter the muscles differentially, whereby they can move the members in all the different ways in which they are capable of being moved[19];" essentially movement of the pineal gland is appended to Descartes' existing motor system as the most upstream component. Sensory perceptions are conveyed to the pineal gland in a precisely analogous fashion; transduction of sensory stimuli pulls the nerves filaments in a particular way which results in particular motion of the animal spirits and then the pineal gland, "which can be moved by them in as many different ways as there are differences capable of being sensed in objects[19]." While this mechanism explains how movements of the pineal gland correspond to sensory and motor function, i.e. "by the mediation of spirits, nerves, and even blood," the pineal gland is not the soul, but far enough downstream that it is a physical entity, and thus this attempted explanation is in fact non-explanatory. This is best summarized in a letter from Princess Elizabeth of Bohemia to Descartes, with whom he corresponded with at length about the implications of his mind-body dualism. Elizabeth writes,

> How can the human soul, which is only a thinking substance, determine the movement of the animal spirits in order to perform a voluntary action? It seems as if every determination of movement results from the following three factors: the pushing of the thing that is moved, the manner in which it is pushed by the body that moves it, and the quality and shape of the latter's surface. The first two presuppose that the bodies touch, while the third presupposes extension. You exclude extension completely from your concept of the soul and, it seems to me, it is incompatible with being an immaterial thing. That is why **I am asking for a more specific definition of the soul than what is provided in your *Metaphysics,*** that is, of the substance of the soul when it is separated from its action of thinking[9]."

This interaction between the non-corporeal soul and the physical body, of which the pineal gland is an intermediary, is Descartes' black box;[***] he has no satisfactory answer for Elizabeth's question. In Cartesian Dualism, which this author joins Baker and Morris in labeling a misrepresentation of Descartes, "thoughts and state of the body are logically independent but causally interrelated[4]." They point out in their introduction that

[***] The term is borrowed from biochemist Michael Behe in his controversial work *Darwin's Black Box*[6] where he questions evolution from a biochemical perspective. Behe uses the term to refer to a subdiscipline that cannot be adequately explained at present, for instance, the cell. While the input and output may be understood, the inner workings of the black box are mysterious until it can be opened and its contents explored. The author feels they analogy is valid here.

"it is notorious that Descartes could give no defensible account of *how* these causal interactions take place[4]," and the proposition is that this is the ceiling of Descartes' capability to explain human mental capacity, and represents the black box in his theory, to be opened when the intellectual and scientific background changed sufficiently to allow it. Skipping momentarily to the conclusion, inside the black box, we find Descartes' *mind* or *soul* is a complex network of a hundred billion neurons and their various connections, continuously firing, changing, and being modulated to constitute our mental world. Further, within this black box are smaller black boxes that remain to be explored. For now though, we return to the nature of this interaction as Descartes understood it by his death in 1650, not to demonstrate it as incorrect, but as incomplete and unsatisfactory.

From *Passions*, "merely by willing something [the soul] makes the little gland to which it is closely joined move in the way required to produce the effect corresponding to this volition[19]." This correspondence has no causal component; "these two streams of events are kept marching in step by God's constant interventions[4]." Such an assertion seems to call into question whether there is even an interaction. Further, in analogy to the non-resemblance between an object in the visual field and its representation on the retina or to an object and the word used to signify it, there is no intrinsic commonality between the movement of the pineal gland and the thought in the soul. Rather,

> "A particular movement of the pineal gland would not be the movement it is if it were not tied to the particular thought which it in fact occasions the soul to have, and a particular volition (of the sort which terminates in an action of the body) would not be the volition it is if it were not tied to the particular movement of the pineal gland which is in fact occasioned by it[4]."

Correlation between thought and movement of the pineal gland is arbitrary, unintelligible, and necessary. The common feature of these two properties, non-causation and non-resemblance, is that they highlight the theocentric aspect of Descartes' explanation of the interaction between mind and body. The strict correlation between a particular thought and the movement of the pineal gland that corresponds to it and effects a particular brain pattern characterized by motion of animal spirits, is 'ordained by God' to ensure and maintain our welfare, and thus could be no other way. It is appropriate that

the physical property of absence of nutrients should be transduced to the brain, perceived as the passion of hunger, move the pineal gland in precisely the appropriate way to dispose the soul to have the volition to eat, that this volition should then incite the motion of the pineal gland that corresponds to inducing feeding behavior, and then that this movement results in the movement of animal spirits that direct the muscles to obtain and ingest food.

In describing this interaction, Descartes refers frequently to 'our Nature as the union of mind and body' and describes this union as an '*ens per se,*' or essential entity rather than an '*ens per accidens,*' or accidental entity. The former phrase emphasizes this property of being ordained by God, hence Baker and Morris use a capital 'N' to highlight this feature in contrast to a simply inherent property or disposition that 'nature' would suggest. The identification of the *Nature* of the entity of the mind-body union as essential rather than accidental reinforces this, and Descartes' use of the two phrases attempts to answer, at least from a theological perspective, his self-proclaimed absurd challenge to conceive of the mind and body as simultaneously distinct and united.[†††]

On New Foundations:

Up to this point the question of Descartes' theology has largely been avoided, and the author does not intend to delve too deeply into his particular religious beliefs. Descartes was a self-professed believer and a practicing Catholic, though his works were placed on the index and there have been various charges questioning his orthodoxy. He declares one of the principle aims of *Meditations* to prove the existence of God, and God does ultimately figure into his explanation of the soul. However the relationship between Descartes' theological and scientific beliefs is less volatile and we can largely consider the former peripheral to our present discussion[‡‡‡]. He does not offer a religious explanation in place of a scientific one *a la* intelligent design versus evolution, nor does he militantly insist on separate domains for religion and science. Rather Descartes' religion emerges as a *de facto* explanation when his science reaches its limit but without

[†††] Perhaps he has the Trinity in mind here.

[‡‡‡] While the interplay between religion and science is very much still a relevant question and could be considered in an equally comprehensive work solely devoted to the subject, this assertion simply maintains that it can largely be separated from the present question.

precluding future compatibility. He seems to subscribe to the school that on the one hand mechanistic scientific explanation does not diminish the power or wonder of God, but on the other his faith leaves open the possibility that some things *may* be unintelligible. Taking this perspective, one can treat God or Nature in Descartes' theory of mind, even where he explicitly figures in as an explanation, as functionally synonymous with another black box, whose contents may very well be spiritual, but may also be neurological[§§§].

Descartes vies with figures like Ptolemy, Franz Joseph Gall, and Jean-Baptiste Lamarck for having amassed the most criticism for his ideas or their implications. He has been criticized by both contemporaries and successors variously for his thermodynamic theory of blood flow, the alleged implication of a homunculus fallacy, and perhaps most poignantly, for his depiction of non-human animals as non-sentient. Descartes maintains that the soul and all of its functions are a distinctly human faculty, and therefore the implication is that in lacking a soul, animals lack consciousness; he is then construed as implying that animals do not feel pain. While he does explicitly deny this charge, his discussion of higher vertebrates proves interesting.

> "even though [animals] have no reason and perhaps no thought either, all the movements of the spirits and the gland that excite the passions in us still exist in them, and serve in them to maintain and strengthen, not the passions as in us, but the nerve and muscle movements that usually accompany them[19]."

Essentially animals share all of the faculties that are not exclusively ascribed to the soul – sensation, imagination, memory, as well as those which do involve the soul, except they necessarily achieve the same ends without the soul. This concession on Descartes' part allows for the interpretation that the soul in humans is peripheral, a construct that serves the sole purpose of endowing humans with a "distinctive power or ability… …in virtue of which some of their actions are subject to moral evaluation[9]," and thus figures more into Descartes' moral philosophy than an explanation of mental phenomena.

This does not constitute a challenge to the sincerity of Descartes' convictions or suppose that he constructed his theory the way it is now being dismantled, layering the soul onto an already viable explanation; in fact this can only be done retroactively for

[§§§] Note that this refers specifically to the hyperbolic doubt → *cogito* → mind/body dualism track that we have followed; there are other aspects of Descartes where theology plays a more critical role and should not be regarded in this way.

reasons which will become evident shortly. In an annotation to his translation of *Passions*, Stephen Voss identifies seven questions that Descartes poses "in comparing higher animals with people: Do animals have life, souls, sensation, passions, thought, reason, the use of language[19]?" and notes that Descartes' answers are complex and often contradictory. Clarke attempts to identify univocal answers to each of these questions in Descartes[9], we makes the same attempt here:

life	yes
souls	no
sensation	yes
passions	in a strict sense no, but a functional equivalent
thought	ambiguous, Descartes supposes not but maintains that this cannot be proven[3]
reason	no
language	no

Clarke launches into a lengthy discussion of the human capacity for language, and attributes a great deal to it in understanding Descartes, though Descartes wrote comparatively very little about language. In contrast to human reason as a "silent and wordless world that would allow undisturbed reliance on intuition[1]," for Clarke "language is essential for metaphysical or abstract thinking," i.e. wordless thought is impossible. Moreover, he distinguishes between human and animal language in that the latter is invariant and reflexive, i.e. animal utterances are the direct and automatic result of external stimuli, whereas human language is rooted in convention and requires rational thought. In the *Fifth Objections*, Pierre Gassendi levels a criticism at Descartes' theory of mind framed from linguistic capacity but extrapolated to other cognitive functions. He challenges "the claim that there is a difference in kind, rather than merely in degree, between the human mind and the corresponding faculties in other animals[9]." He maintains that "although human being are foremost among animals, they still belong to the same class[9]," i.e. the question that we have already entertained as to whether there is a

fundamental discrete difference between humans and other animals or whether the distinction ranges along a continuum.

In his response to Gassendi in 1637, Descartes writes with no way of having anticipated how drastically the rules of the game would change two hundred and twelve years later when a British naturalist named Charles Darwin would publish *The Origin of Species*. By the old rules, Descartes' explanation is plausible; after 1859, "*Nothing in Biology Makes Sense Except in the Light of Evolution*[22]." Theodosius Dobzhansky concludes his paper by that name, "evolution is a light which illuminates all facts, a trajectory which all lines of thought must follow.[22]" Considering Descartes in the context of evolution, obviously a necessary anachronism, though one Descartes would readily concede to, renders the supposition of a fundamental distinction between humans and animals absurd, and in Descartes' own words, "undermining the foundations will cause whatever has been built upon them to crumble of its own accord[17]."

The other major charge leveled at Descartes is that the mind-body dichotomy that emerges in his works leads to what has come to be known as Cartesian Dualism. While nominally associated with him, this system represents four centuries of distortion, and the thesis of Baker and Morris' *Descartes' Dualism* is that Cartesian Dualism, or what they label to the Cartesian Myth, would be scarcely recognizable to René Descartes. Desmond Clarke, Dennis Sepper, and this author join this consensus, and while the nuances of what each proposes as an alternative differ, this author adds his voice to the univocal attestation that Cartesian Dualism is not a necessary, preferential or perhaps even possible interpretation of the primary Cartesian texts.

In the second chapter of *Descartes' Dualism*, the name for their proposed reinterpretation, Baker and Morris identify four tenets of classical Cartesian Dualism[4]:

1. two distinct worlds, one consisting of physical objects and the other of mental objects
2. physical objects characterized by mechanistic operation (clockwork) and mental objects by consciousness
3. objective accessibility of physical objects by the senses in contrast to private, privileged, and infallible knowledge of mental objects
4. causal interactionism between physical and mental objects

A battery of unpleasant implications follows. The 'two-world view' creates an unnecessary pleonastic system akin to the Scholastic tradition that Descartes sought to break with. A corollary of the 'consciousness and clockwork' dichotomy is the *expansion thesis*, taken from Descartes' definition of a thinking thing as 'thing that doubts, understands, affirms, denies, wills, refuses, and also that imagines and senses,' which expands the functions of the mind or soul at the expense of those of the body, the so called *contraction thesis*, moving the boundary between mind and body from that between 'intellect' and 'sense' to between 'consciousness' and 'clockwork.' Cartesian introspection, our infallible knowledge of our own mind, is described in analogy with sense perception, and the plethora of difficulties with causal interactionism have already been explored at length. Baker and Morris summarize Cartesian Dualism as "a picture of two worlds, one of physical objects known through sense-perception and one of mental objects known through infallible introspection which (mysteriously!) causally interact with each other," and then assert that this is "entirely *at odds* with the very framework of Descartes' thinking[4]."

In what they name 'Descartes' dualism' to contrast it with Cartesian dualism, Baker and Morris enumerate a parallel set of maxims that they believe more accurately reflects Descartes' own views as expressed in his texts:

1. there are two and only two kinds of substances, corporeal things and thinking things
2. the essence of the mind is thought, the essence of the body is extension
3. bodies and their properties are objects of sense-perception; minds and their properties cannot be
4. human mind and body are 'substantially united' but interaction between mind and body is 'rationally unintelligible'

Their system is certainly truer to the texts that we have surveyed thus far, though at the outset a claim was made that the author would take license to deviate where appropriate from the explicit statements of seventeenth century Descartes to the corresponding relevant ideological homologs. This license is applied here to deemphasize the metaphysical distinction between mind and body that comprises the first two of these

maxims, and to highlight a salient feature of Baker and Morris' version of Descartes' dualism.

A feature of the Aristotelian theory of mind that predominated prior to Descartes was the classification of predicates into four types that corresponded to the type of soul that was required. The simplest were merely descriptive predicates, C-predicates, that could be applied to corporeal objects; "the car **was red**." Ascending the hierarchy, the next step requires a vegetative soul, ascribed to plants, animals, essentially anything organic; "the bacteria **was dividing**." Next are S-predicates, which required a 'sensitive[****] soul;' "the dog **sees light**." Finally, the highest class, the R-predicates, requiring a rational soul, are ascribed only to humans; "the reader **judges that this exercise is slightly ridiculous**." Descartes dispenses with the need for a separate class of S-predicates by pointing out their inherent ambiguity; a member of this class must correspond to a corporeal body or to a rational soul, but not simultaneously to both; "the dog $sees_1$ light is of the former" while "the dog $sees_2$ light" is of the latter. This is not just an exercise in semantics, but rather now only the precisely defined two-subscript verbs are associated with mental substances; everything else is associated with corporeal bodies, and the construct of the 'sensitive soul' drops out.

This narrowing of the domain of the soul is quintessential feature of Baker and Morris' version of Descartes' dualism, and also the most important observation to be gleaned from our distillation of the Cartesian texts. In stark contrast to the expansion thesis, and the accusation that Descartes has perpetuated a non-physical entity as an essential feature of understanding human mental capacity, this interpretation suggests precisely the opposite. Instead, Descartes attempted to carve out a specific niche for the soul that contained only the functions absolutely unique to it, intellect and volition, or moral agency, while simultaneously expanding what could be accounted for mechanically. Consistent with the trend noted through his physiology and natural philosophy, Descartes sought to genuinely and adequately explain as many phenomena as possible. Baker and Morris note that

> "His vision (or programme) of extending the scope of mechanical explanation has affinities with the modern prejudice that physical explanations underlie all

[****] Literally, capable of sensing

> phenomena (that quantum mechanics is the least common denominator of all scientific explanations, and that other explanations, for example, in chemistry or biology, are simply place-holders for a complete quantum-mechanical explanation to be worked out in the scientific millennium)[4]."

By analogy, Descartes' theory of mind as it stands is a placeholder for a complete neuroscientific explanation. While Baker and Morris qualify this, identifying an "absolute or metaphysical limit on the possibility of scientific explanation[4]," Clarke, with whom this author agrees here, is more ambitious, maintaining that "the Cartesian theory of mind is a first rather bold step in the direction of removing souls completely from explanations of human behaviour, because they are mere non-explanatory redescriptions of the phenomena to be explained[9]." In the first half of the seventeenth century, Descartes could not satisfactorily eliminate the construct of the soul completely, and this likely was not his explicit goal, but he did succeed in attributing as many functions as possible to the physiological brain. The soul remained as a black box to house the aspects of human mental function that were inaccessible in his time. However we now have the tools to open this black box and continue the process that he started, rendering the non-physical mind purely vestigial. To the capacities he left to the soul, intellect and volition or moral agency, we now apply the modern term 'executive function.' The present goal is to take Descartes' work a step further, and assert that the totality of our mental and cognitive capacities can be, and necessarily must be, explained in a single domain, thus making the mind-brain distinction superfluous.

Inside the Black Box:

Having qualified the nature of the mind-brain dichotomy as it emerges from Cartesian philosophy and liberated Descartes from the gross distortions that have attached themselves to his name over the past four centuries, the present investigation now intends to reconcile this version of Descartes with the scientific landscape of the twenty-first century. The preceding analysis identified a *restricted*, contrary to the presupposed *expanded* version suggested by Cartesian dualism, domain belonging exclusively to the non-corporeal soul, namely intellect and volition. The intention is to demonstrate that all aspects of this greatly reduced domain can now, though perhaps they

could not in 1650, be explained entirely in terms of a physiological brain and to render the hypothesis of a disembodied soul unnecessary.

The author believes that had Descartes had at his disposal the theory of evolution and neuron doctrine he would have necessarily arrived at the same conclusion, namely that the soul has become pleonastic. Thus the present discussion does not intend to suggest Descartes' theory is antiquated and obsolete but rather that his ideas are largely compatible with modern neuroscience and that the past four centuries have provided the tools to open the black box encasing the highest levels of human cognitive function. As will become apparent, the black box of the human mind[*] is akin to a set of Russian nesting dolls; opening one black box yields others within it. However, a feature of the landscape as it stands today is that this analogy holds completely; there is necessarily a solid doll at the center and not an infinite regress. One can now assert the plausibility of explaining the entirety of the workings of the human mind entirely in terms of neurons and their interactions. This is not meant to suggest that 21st century neuroscientists possess a complete understanding of the human brain, but that one is on the horizon, refuting the suggestion that there is an "absolute or metaphysical limit on the possibility of scientific explanation[4]" with regard to the human mind.

Before embarking on this journey, it proves useful to highlight some of the fundamental changes in scientific thought between Descartes' time and ours that warrant this paradigm shift and allow one to dismiss the idea of a non-corporeal soul because it is no longer necessary, useful, or even plausible. We have already discussed evolution, but it is worth emphasizing again here. The Dobzhansky paper captures the most essential feature of modern biology, science, and perhaps even thought in general, in its title, *Nothing in Biology Makes Sense Except in the Light of Evolution*[22]. Once one accepts the evolution of biological systems, on global, species, and biochemical levels, everything must subsequently be understood within this framework. There can be no fundamental distinction between humans and other animals; we must necessarily be viewed along a continuum. There is no teleology in evolution; structures, systems, and behavior patterns evolved out of precursors because they possessed some feature that conferred a survival advantage on individuals within a population. Though not the focus of the present

[**] Now taken to reside entirely within the physical brain

investigation, the evolutionary context is a critical feature that one must remain cognizant of throughout the discussion. In *The Growth of Biological Thought*, Ernst Mayr notes that a "difficulty for the historian is posed by most scientists' unawareness of their own framework of ideas. They rarely articulate – if they think about it at all – what truths or concepts they accept without question and what others they totally reject[30]." The author finds this an appropriate caution and wishes to avoid this pitfall in the present analysis. To allow too much into the former category is dangerous, and rather than amass a list of axioms, the author simply intends to note that the 'framework of ideas' of the twenty-first century is necessarily different from that of the seventeenth. In addition to evolution, the other fundamental concept that transformed the landscape is cell theory, namely that living organisms are composed entirely of cells and that cells arise only from other cells.

A Brief History of Post-Cartesian Neuroscience[†]:

It proves useful to the present discussion to concisely review the development of neuroscience from its infancy in Descartes' time to the present. This is by no means an exhaustive treatment but merely intended to set the context for our investigation of neural phenomena and also to highlight what was not known in Descartes' time.

The nervous system consists of the brain, the spinal cord, and afferent (sensory) and efferent (motor) peripheral nerves. Brain structures and nerve fibers are composed of cells called neurons, which are excitable cells that communicate via electrochemical signaling. This occurs via something called an action potential, whereby an electrical impulse resulting in depolarization of the cell membrane is conducted from the cell body in one direction down the cell's axon, a long projection for the purpose of carrying these signals to the target neuron. At the axon terminus, chemical messengers called neurotransmitters are released into the synaptic cleft, the small gap between neurons, which bind to receptors on the target neuron. Neural connections can either be excitatory, in which they depolarize the subsequent neuron and increase its likelihood of firing an action potential, or inhibitory, in which they hyperpolarize it and decrease this likelihood. Action potentials are all-or-none, meaning that a cell either fires one or does

[†] The Hawking reference is intentional[28]

not, and they contain no inherent qualitative information. Rather information is conveyed by the rate of firing, the population of neurons firing, and the synchrony or lack thereof.

There are a number of discrete structures within the brain, each with particular functions. It is not relevant to the present discussion to catalog these here or enter into a detailed discussion of neuroanatomy[‡]; instead the author intends simply to highlight a few historical and practical aspects of the idea of localization of brain function. The scientific community has vacillated back and forth on this question in a largely reactionary fashion. On the one extreme were the phrenologists who created detailed maps of the brain that contained a center for every conceivable intellectual or moral faculty. A person's temperament could be inferred from the shape of their skull because these faculties were correlated with the size of corresponding brain centers, and the iron bar that destroyed Phineas Gage's frontal lobes[§] passed somewhere "in the neighborhood of Benevolence and the front part of Veneration[12]." At the opposite extreme is the denial of any localization of function; according to Flourens, "The brain secretes thought as the liver secretes bile[37]." Obviously the truth lies somewhere between these two radical hypotheses. The phrenologists did contribute the concept of brain specialization, but the notion of brain centers proves too restrictive. It is readily apparent that at a tissue level the brain is heterogeneous; Brodmann identified forty-six areas based on cytoarchitectonic differences, which unsurprisingly correspond to functional divisions, for example, Brodmann's area 17 is primary visual cortex.

Thus different brain processes occur at different locations in the brain. Regions of the brainstem control vegetative and homeostatic operations. There are specific regions, both cortical and subcortical, devoted to processing specific modes of sensory information. For example, auditory information is processed in the cortex of the temporal lobes and the medial geniculate nuclei of the thalamus. Similarly there are regions involved in motor function that project from motor cortex to the spinal cord and then to the skeletal muscles. On the one hand, this localization of function can be highly specific; for instance lesions to Broca's area, a tiny region of the left frontal lobe, selectively disrupts expressive language. We will return to this theme shortly. However

---

[‡] Although familiarity with basic neuroanatomy is certainly advantageous to the discussions that follow

[§] Discussed at length below

the nervous system is more complex than this and most of its faculties are far more diffuse, and represented in neural networks that span many brain regions rather than residing in specific loci.

Francis Crick emphasizes the appropriateness of this reductionist approach to neuroscience, namely that a complex system can be explained entirely in terms of its component parts and their interactions. This structural reductionism is evident at the juncture of biology, physics, and chemistry, where one explains organisms in terms of cells, cells in terms of molecules, molecules in terms of atoms, and atoms in terms of quantum mechanics. The interactions between the parts of a system are as essential to its understanding as the parts themselves; Crick offers the benzene molecule as an example. Six atoms of carbon and twelve atoms of hydrogen tell part of the story, but it is necessary to consider the diffuse electrons and aromaticity to explain the reactivity and physical structure of the molecule. This interaction is perhaps even more important when one considers the nervous system. While neurons do vary, in shape, size, number and distribution of neurites, type of neurotransmitter used, degree of myelination of their axons, and various other properties, they are on the other hand more similar than they are different, consisting of dendrites, a soma that contains all the features of any other cell, and a single axon that carries a single type of signal. To explain the capacities of even the most basic brain, nevermind the intricacies of the human mind, solely in terms of individual neurons would be a daunting task; it is the nature of the connections between these hundred billion cells from which we derive the emergent behavior of the nervous system.

Localization of brain function, to the extent that it does occur, has provided one of the most valuable tools for studying neuroscience. If damage to brain region A results in disturbance in behavior or function X, then one can conclude that A is essential for X, or that A to some extent occurs in X. Moreover since by definition A is an element of the set that comprises the physical brain, then X can in turn be attributed to the physical brain. Antonio Damasio further describes this experimental neuropsychological approach for “finding systematic correlations between damage at given brain sites and disturbances of behavior and cognition” in terms of finding double dissociations, whereby “damage at site A causes disturbance X but not disturbance Y, while damage at site B causes

disturbance Y but not disturbance X[12]." The next step is to use these observations to formulate hypotheses about function in a normal brain, the goal being to "explain how certain cognitive operations and their components relate to neural systems and their components," not, Damasio cautions, about "finding the brain 'localization' for a 'symptom' or 'syndrome.'" Thus in our present investigation, though it may be an attractive hypothesis, we have no need for a conclusion of the form *the will is located in the anterior cingulate gyrus*, but merely one of the form *selective brain lesions can perturb functions associated with the will, therefore it is likely that this faculty can be explained in terms of the physical brain*.

Lesion studies provide the primary means for this type of investigation, and have yielded a wealth of information. In humans we consider unintended brain lesions resulting from injury, stroke, or neurodevelopmental disorder whereby a specific brain region is damaged, or deliberate surgical lesions performed, usually to correct a more debilitating problem such as epilepsy. However nature pays no attention to cytoarchitecture, and these lesions are often crude. Animal experiments allow us to use a finer point brush and perform deliberate selective lesions that would be unethical in humans. Complementing the information learned from lesion studies is a more transient and less invasive but perhaps equally informative tool – neuroimaging. Using positron emission tomography (PET) and functional magnetic resonance imaging (fMRI), it is possible to view relative brain activity measured by blood flow and glucose uptake in real time to determine which brain regions are most active during particular brain processes. Combined, these two approaches give us a powerful window into the inner workings of the nervous system that was unavailable to Descartes, and the hope is that from this better vantage point, we can illuminate a physiological explanation for those faculties that remained inside Descartes' black box.

Having established sufficient Cartesian and neuroscientific context, we now allow these two tracts to intersect. Essentially the programme is to catalog a series of the most fascinating aspects of the human *mind* that can be explained in terms of the physiological brain. Structurally this section loosely recapitulates the previous; we consider approximately the same topics in the same order, beginning with sensation and ending with the will. The intention is that with regard to each capacity that Descartes left to the

soul, the exponential increase in explanatory power over the past four hundred years will provide a striking enough contrast for the reader to conclude that a disembodied soul is no longer a necessary hypothesis. By considering some of the most subtle and 'higher-order' phenomena within each of these domains, the author intends to provide a counterpoint to Descartes' supposition that for example, most sensory processing occurs within the brain but at its highest levels requires the soul. The process represents a continuation of Descartes narrowing the domain of the soul, with the additional goal of rendering it a potentially empty set and thus a null hypothesis that can be rejected.

Sensory Phenomena:

The account of sensation that emerges from Descartes is ambiguous; "[sensing,] precisely so taken is nothing other than thinking[17]" though it "surely does not take place without a body[17]," yet simultaneously, "what I thought I had seen with my eyes, I actually grasped solely with the faculty of judgment, which is in my mind[17]." Even if one accepts the previous arguments whereby sensation was largely allocated to the physical brain even in Descartes' system, he does still to some extent attribute sensation$_2$ to the disembodied mind, and thus this proves a useful place to begin. Descartes himself discusses sensation at length, and a great deal of work in modern neuroscience has been done with regard to sensation, because of its relative approachability and the comparative ease with which it is understood. Francis Crick identifies several reasons for using sensation, particularly visual perception, which is where we will begin, though other sensory systems will be considered, as a springboard and archetypical example of brain function[11]. He notes that our sensory world is vivid and rich in information, sensory input is structured but easy to control, and it is conducive to animal experiments. While this concreteness establishes sensation as a suitable entry point into our investigation, it is far from trivial, and the intricacies of the human sensory world, and their capability to be explained in terms of the brain, will readily become manifest.

The Human Visual System:

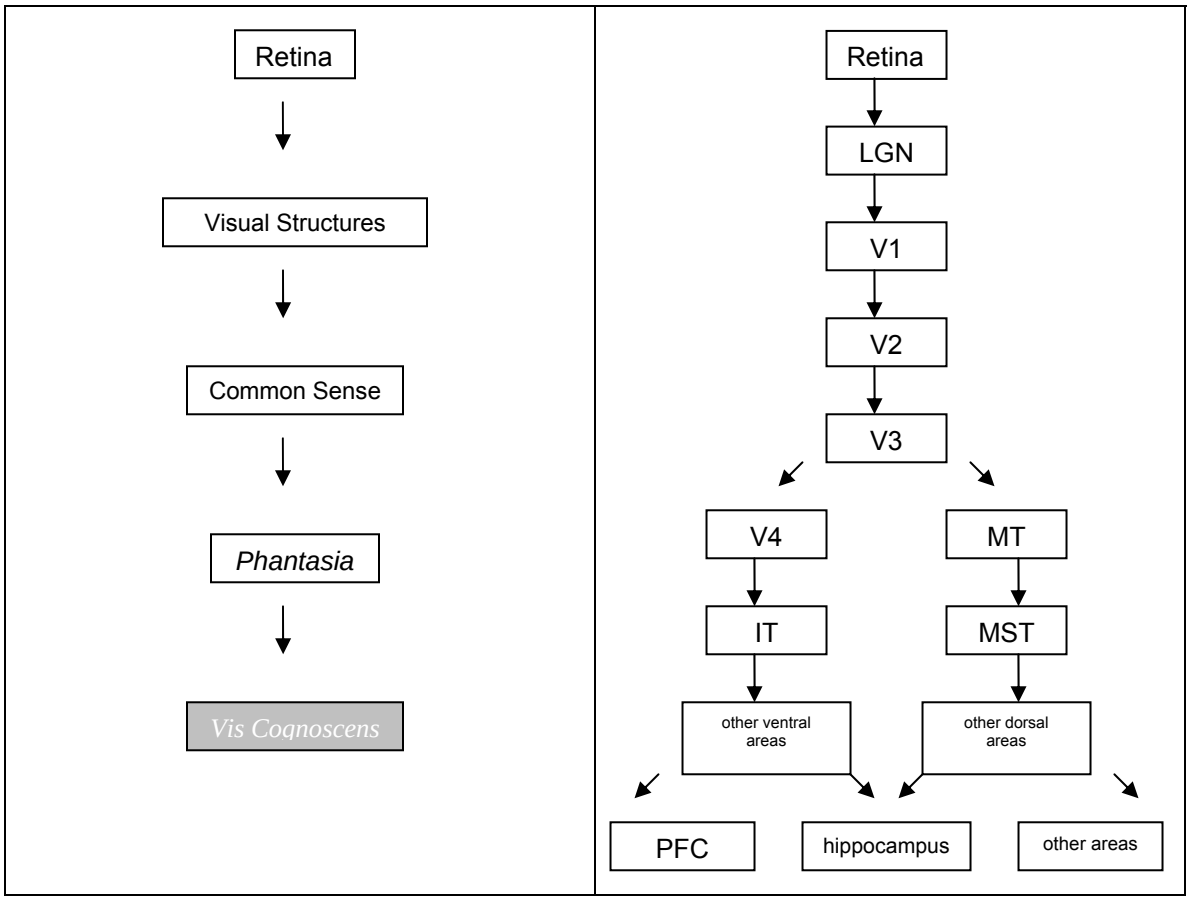
Descartes gave a comprehensive account of his understanding of the human visual system, and as was noted above, his explanations were remarkably accurate, especially

considering the context. Both for comparison and also as background, we present an account of the human visual system as it is understood today and juxtapose Descartes' for the purpose of establishing the functional homologs between the two models in order to circumscribe those faculties most likely to be deemed extra-physiological. Subsequent examples are spared such painstaking detail, but the author deems it worthwhile to explicate the process once.

Visual information in the environment, i.e. light reflected from physical objects, impinges on the sense organs, namely the eyes. Some physical optical processes occur that are not especially relevant; our neural model begins with the absorption of photons, or particles of light, by the pigment within the photoreceptor cells of the retina. The physical stimulus, light, is transduced into a neural signal by these cells. Photons absorbed by the photopigment in the photoreceptors cause a conformational change in photopigment that signals, through a biochemical cascade, the closing of specific ion channels that elicits a change in membrane potential. In this way, physical energy in the form of photons is converted into an electrical signal. Some neural processing about the physical stimulus occurs in the retina itself, and eventually information about the visual information in the environment, now represented entirely in the form of an electrical signal, is further transformed as it is carried to subsequent nervous system structures. The retina projects primarily to a subcortical structure called the thalamus, specifically the lateral geniculate nucleus (LGN). There are a few important features about this projection to point out. First, there is a retinotopic map, or representation of the visual field on the retina, which is repeated in subsequent structures. Equally important, information remains segregated; different neurons fire differentially in response to different features of the stimuli – color, orientation, motion, location, and these project to cells along segregated pathways in which all neurons are sensitive to the same types of features. An emergent property from these two is that certain fibers from each eye cross the midline, such that information in the right visual hemifield is processed by the left LGN and information in the left visual hemifield is processed by the right LGN. Finally, the LGN is not the only target of the retina; there are other projections involved in orientation, circadian rhythms, and movement of the ocular muscles.

The LGN then projects to primary visual cortex, a region of the occipital lobe. This region in turn projects to extrastriate visual areas, a series of regions that process increasingly complex features of the visual stimulus. The immediate targets of primary visual cortex are other occipital areas, V2 and V3, and then the information is directed along two parallel streams that maintain the segregation inherent in the pathway since the retina. The dorsal stream projects through the parietal lobe and processes features concerned with visual motion. The ventral stream projects through the temporal lobe and is concerned largely with object recognition. High level processing occurs here along each of these streams; for example, selective lesions to certain parietal regions can destroy solely one's ability to perceive motion, and there is region of the inferotemporal lobe that may be involved exclusively in face recognition. These two streams then project to a wide range of cortical and subcortical targets whereby visual information is integrated with information from other sensory modalities, stored in memory, used to influence decision making, volitional movement, and a multitude of other responses. An important feature of this pathway is that it is not unidirectional; while information is largely segregated, there is crosstalk between these semi-independent pathways at all levels, between cortical layers, between the dorsal and ventral streams, and between this pathway and extrastriate visual pathways. Moreover, a significant amount of feedback to earlier stages of processing occurs as well.

Consider this pathway juxtaposed with Descartes' below. This diagram is an oversimplification, especially on the right side. It neglects to show the bidirectionality, omits some projects, and does not include much of the crosstalk or parallel pathways. The intent is not to faithfully depict the intricacies of the visual system, but rather to provide a schematic diagram for comparative purposes:



Note that both models begin with the retina, and the subsequent brain structures largely correspond. Descartes identifies specific brain structures associated with vision, which we will equate with the LGN and primary visual cortex. The *common sense* corresponds to multisensory integration, which actually occurs somewhat further beyond the occipital lobe. Descartes' *phantasia*, or imagination, can be equated with consciousness, or visual cortex, both primary and extrastriate (V1, V2, V3, V4). Finally, the *vis cognoscens*, or 'force with which we properly know things,' corresponds to these highest levels of visual processing, which from the diagram necessarily overlaps with conscious perception, but is also downstream of the most basic realization of this. The intention here is to demonstrate that the most salient features of human visual perception can be explained entirely in terms of these most downstream structures, and therefore equating Descartes' *vis cognoscens* with the set comprised entirely of these structures is a reasonable conclusion.

Selective lesions to areas of the brain involved in visual processing can produce highly specific deficits. The retinotopic organization of the visual system predicts that lesions to areas of the retina, LGN, or striate cortex, will result in a scotoma, or blind spot in the part of the visual field corresponding to the lesioned tissue. For instance, one would expect, and this is indeed the case*, that complete ablation of primary visual cortex on the right side would result in complete loss of conscious vision in the left visual hemifield. As one progresses further downstream, the results of lesions become more interesting. Damage to regions along the ventral stream can result in the selective loss of color vision, a condition known as cerebral achromatopsia†. The most striking feature of cerebral achromatopsia is that subjects, who previously had normal color vision, not only completely lose the ability to see in color, but also can no longer imagine colors that were previously known to them. An equally specific disturbance to visual perception is a condition known as prosopagnosia, where one loses the ability to recognize faces. Affected persons retain the ability to recognize a face as a face, as well as to perceive its individual features – they can describe the eyes, the nose, the mouth, but somehow lose the ability to perceive the face as a gestalt or to identify to whom a particular face belongs, including even one's own face. Damage to areas in the dorsal stream interferes with one's ability to perceive movement, which is difficult to conceive. Affected persons see still images, an object in one place, and then in another, but somehow are unable to perceive fluidity of motion and see things perhaps as if illuminated by a strobe light. One can imagine the difficulty such a person would have in crossing the street or filling a cup of tea, both of which are noted in Francis Crick's account of a case study by Joseph Zihl[11]. Crick draws the comparison to our failure to perceive the hour hand of a clock as moving, but on an accelerated timescale.

Blindsight:

Perhaps the most intriguing visual anomaly is a phenomenon known as blindsight, which is defined by the *Oxford Concise Medical Dictionary* as "a condition in which the

---

* Albeit with some remarkable qualifications that form the basis for the discussion of blindsight that follows

† Oliver Sacks describes a very interesting case of cerebral achromatopsia in *The Colorblind Painter* in his book *Anthropologist on Mars*[37]

sufferer responds to visual stimuli without consciously perceiving them[31]," or alternatively by Larry Weiskrantz, the Oxford psychologist who has most extensively studied the phenomenon and given it its name, as "visual capacity in a field defect in the absence of acknowledged awareness[41]." As promised, the statement made above, that 'complete ablation of primary visual cortex on the right side would result in complete loss vision in the left visual hemifield,' is now slightly revised, and we append the familiar 'two' subscript to 'vision,' to emphasize that such a lesion completely destroys conscious perception of vision. In blindsight however, some residual visual capacity is retained in spite of this complete perceptual blindness. Subjects adamantly deny any vision in the blind part of the visual field, yet there is a discrepancy between this verbal denial and performance on tasks that assess visual function.

The retinogeniculocortical visual projection described above, where visual information flows from the retina to the LGN to primary visual cortex, is the primary pathway mediating vision, and perhaps the only one mediating conscious vision, but it is not the only visual pathway. Blindsight is characterized by lesions to primary visual cortex; pregeniculate lesions are accompanied by minimal, if any, residual visual capacity[41]. Primary visual cortex is required for conscious perception of vision, but a parallel pathway through the midbrain, where retinal neurons project to a structure called the superior colliculus, which in turn has targets in the pulvinar nucleus of the thalamus, that projects to extrastriate visual cortex, has been implicated in the preservation of some of the visual capacities retained in blindsight patients. Because cortical blindness frequently results from occipital damage caused by trauma or stroke, it is often unilateral, resulting in a hemianopia, or loss of vision in only one visual hemifield. Bilateral striate cortex lesions are rare, and most blindsight patients therefore have a visual field divided into a region where they have normal conscious vision and a region in which they are completely blind.

The historical context in which Weiskrantz began his investigation of blindsight prompts a return to the question of whether there is a fundamental difference between humans and other animals. While the primary visual pathway had been identified, a growing number of researchers began to confront seemingly paradoxical observations regarding occipital lesions in human versus nonhuman subjects[41], calling into question

whether striate cortex was indeed essential for vision. Complete lesions of striate cortex in monkeys sometimes seemed to have minimal effects on vision, whereas in humans scotomas were observed with no residual visual capacity[41]. The resolution of this paradox is that a phenomenon like blindsight, and its consequent implications, is discovered when surprising observations manifest under specific methods of testing. A discrepancy between what a subject $sees_1$ and $sees_2$ can only be identified if the subject is able to report what he or she $sees_2$, which is facilitated by a human subject, though the inability of a nonhuman primate to convey with ease whether or not it can see_2 does not preclude this capacity and require one to suppose a fundamental difference.

The textbook blindsight case is a patient known as D.B., whom Weiskrantz began studying in 1974. D.B. had an "arteriovenous malformation at the pole of the right occipital lobe[41]," which resulted in severely debilitating headaches that increased in frequency and severity. The region was surgically excised, resulting in removal of much of his right striate cortex and the expected scotoma in his left visual hemifield. The remarkable feature of his case however, is that "D.B. appeared to be able to locate objects in his supposed blind field much more skillfully than one might have expected. For example, even though D.B. could not see one's outstretched hand, he seemed to reach for it accurately[41]." Weiskrantz's initial informal assessment of D.B.'s residual visual capacity involved presenting a stick in either a horizontal or vertical orientation to the blind part of D.B.'s visual field, and asking him to guess its orientation. D.B. performed nearly perfectly, but insisted that he did not see the stick at all, and was merely guessing.

The initial stick experiment forms the model for the subsequent tests that Weiskrantz performed to attempt to qualify what visual capacity D.B. retained. Stimuli were presented to the blind part of his visual field, and D.B. was asked to report on various qualities of the stimuli based solely on guessing. General features of the experiments were controls for ambient lighting, and variables included visual angle (distance from the fovea), illumination, orientation, and size of stimuli. Eye fixation, or D.B.'s ability to direct his eyes to a light shone in his blind field was assessed with moderate performance. His ability to point to the location of a stimulus was better, with performance significantly above chance, often around 80-90%. Similar results were observed for presence/absence experiments and identification of orientation. Visual

acuity was assessed by presenting a grating and determining whether he was able to identify it as such. Other experiments assessed his ability to detect movement, which was a function of velocity, and his ability to discriminate forms; he could readily distinguish between X's and O's, but not between an X and a triangle and only above a minimum size that was larger than that required for the same discrimination in his intact visual field. D.B. was also able to identify the direction of contrast, i.e. a black spot on a grey background versus a white spot on the same background.

The contrast between his performance on these discrimination tasks, which was at times near perfect, and his persistent assertion that he was merely guessing, is striking. When questioned about the tasks, D.B. "stressed that he saw nothing in the sense of 'seeing,' that typically he was guessing, and was at a loss for words to describe any conscious perception[41]." He was continually surprised at how well he performed on the tasks. Moreover, as the testing progressed over several years and the tasks became more complex or he became more experienced with them, some subtleties emerged in his reporting, though it never deviated from the general trend of near complete denial of conscious perception and insistence that he was largely guessing. Psychometric tests like these are often conducted only for a relatively short duration at a time because of diminishing performance with the onset of fatigue in the subject. Weiskrantz once expressed this concern, to which D.B. replied, "I am not tired; I haven't been doing anything[41]."

D.B. never reported seeing$_2$ anything in his blind field, but when stimuli were presented with high contrast, he said on one account, "I felt as if something was coming up to me. But I didn't *see* anything[41]." When the contrast was reduced, he described his perception as "Nothing at all. Just guessing. No feelings of anything. No idea at all[41]." Similar variation was observed with respect to moving stimuli; Weiskrantz actually was able to identify a threshold at which D.B. asserted, "I could *feel* the movement, and I was absolutely sure of it[41]," but again he insisted that he could not see anything. Sometimes he described an experience of 'waves' in his blind visual field, which had various differential properties that sometimes formed the basis for his discrimination between stimuli, though again he insisted that he was not *seeing* and that the waves are "unlike anything in normal visual experience, and for which precise words seem to be lacking[41]."

Further, there was no correlation whatsoever with either his identification of some form of percept or his report of how well he believed he performed on a particular task and his actual performance. When D.B. claimed he felt something, or identified the experience of waves, he was no more inclined to correctly identify features of stimuli than when he maintained he was just guessing. Similarly, he would often predict his performance was roughly chance when it was near perfect or assert with confidence that he may have missed one or two trials when his performance was in fact no better than chance.

Weiskrantz's extended study of D.B. has revealed a great deal about the phenomenon of blindsight, though a few observations from other cases are worth noting. A study by Zihl identified both improvement with practice in detection of the presence or absence of a stimulus, and also differential results based on response mode. Subjects performed worse when required to indicate an affirmative response by verbally stating 'yes' as opposed to merely initiating a voluntary eye-blink or pressing a button[41]. Another study by Marzi et al. exploits the observation that a normal subject's reaction time is faster to two simultaneous flashes than a single flash. In a clever modification to this experiment, the reaction times of a blindsighted subject were compared from a single flash presented in the intact visual field to two flashes, one of which was presented in the intact visual field and one in the blind visual field. The shorter reaction time observed in the second case suggests that the flash in the blind visual field contributed to the performance. In another experiment, subjects were able to differentiate between a full circle presented such that half of it appeared in the blind visual field and a semicircle presented at the interface of the intact and blind fields; the subjects completed the figure only when it actually was appropriate to. Finally, a study by Marcel identifies some degree of semantic processing in the blind visual field. He flashed a word in the blind field, the presence or absence of subjects could not detect better than chance, "but the interpretation of an auditorily presented polysemous word (bank) nevertheless could be biased depending on the meaning of the unseen word (river or money)[41]."

Beatrice de Gelder presents a final case for consideration, a rare case of complete bilateral loss of striate cortex[23]. The patient, T.N. suffered two strokes in succession, and Gelder et al. note via imaging studies no remaining active visual cortex. T.N. had worse

residual function than D.B., but exhibited "sophisticated visuo-spatial skills in the absence of perceptual awareness[23];" he could "successfully navigate down the extent of a long corridor in which various barriers were placed."

Not all of these capacities have been assessed in all subjects; it is highly unlikely they all retain the same residual visual capacities, but the complex and diverse functions that are retained reveal a remarkable dissociation between visual processing and conscious perception of such. Weiskrantz summarizes the hallmark feature of blindsight, "where a patient shows a loss of awareness of any vision, or at best only a very degraded awareness, while at the same time displaying a sensitive capacity for detection and for performing certain discriminations that are independent of the requirement for acknowledging awareness" and emphasizes the "disjunction between their verbal reports and their discriminative capacity[41]." The inherent dependence of this phenomenon on subjective reporting has an interesting implication. Weiskrantz notes that the distinction between 'seeing' and 'not seeing' and 'guessing x' and 'guessing non-x' corresponds to the same psychophysical procedure, simply with different response categories. However, comparison of the same task performed by the same subject in seeing versus non-seeing fields illuminates the nature of visual awareness:

> "With supra-threshold stimuli for orientation, movement, or detection, we may find a virtually perfect performance, and yet the subject is adamant that they are quite different in quality: he is patently aware of one but not the other. If pressed, he does not say 'I can make a verbal response to one but not the other;' he insists that they give rise to different experiences[41]"

The ability to make a verbal response is a window into human consciousness, but it is not its only feature. These cases reveal that something as specific, complex, and intimate to the human mind as conscious perception of visual stimuli can be selectively destroyed with the preservation of nearly every other faculty.

Amusia:

Humans have an interesting and complex relationship with music. Darwin writes in *The Descent of Man*, "As neither the enjoyment nor the capacity of producing musical notes are faculties of the least use to man… they must be ranked among the most

mysterious with which he is endowed[38]." Sidestepping the evolutionary question, one can still recognize the extent to which humans are a musical species, and perhaps with the exception of songbirds, the only such. Descartes wrote little about music, but its inherent status as a higher level of sensory perception, possible exclusivity to humans, and its aesthetic and emotional qualities render music a candidate for a feature of the human mind that could reside near the hypothetical body-soul interface that we are attempting to eliminate.

As a sensory percept, music is a complex entity, consisting of both basic auditory properties – intensity and frequency, and also those particular to music – tone, timbre, melody, and rhythm. One would expect focal lesions of the auditory cortex to produce particular predictable deficits, irresponsiveness to a particular frequency or complete tone deafness are both conceivable, for instance. However the more interesting phenomenon is amusia, where "tones are not recognized as tones, and music, therefore, is not experienced as music[38]."

Oliver Sacks[‡] describes patient D.L., a seventy-six year old woman with congenital amusia, the dissociated inability to perceive music as music[38]. The more basic elements of her auditory system are intact; D.L. has no hearing deficit or problem with expressive or receptive language. She was not tone deaf nor had any difficulty imitating a rhythm; in fact she was a tap dancer as a young girl and now participates in rhythmic aerobics.

However D.L. simply cannot perceive music as music. She recalls that in kindergarten children were asked to sing their names, and that she could not do this nor understand what everyone else was doing. She was unable to identify songs; even as a teacher she could not recognize "Happy Birthday" despite that fact that she would play it at least thirty times a year. When asked to describe what other people hear as music, D.L. responded, "If you were in my kitchen and threw all the pots and pans on the floor, that's what I hear[38]!" Another amusic likened his experience to 'a screeching car,' and

[‡] Sacks actually describes a number of phenomena relating to music in *Musciophilia*[38], including musical hallucinations, musical seizures, various types of amusia, savant-like musical talents, absolute pitch, musical synesthesia, and the relationship of music with various disorders including amnesia, Tourette's, Parkinson's disease, Williams syndrome, and dementia.

another maintained that the extent of his discriminatory capacity was that "he could say only that [a particular musical piece] was "The Marseillaise" or that it was not[38]."

This perplexing condition, the selective dissociated loss of the capacity to perceive music, provides another window into the types of processes that must occur in the brain and further diminishes the likelihood of an entity that functions outside of this system.

Somatosensory Anomalies:

The somatosensory system, through which we perceive sensations of touch, pain, temperature, body position, and body movement is unique in that these sensations are "localized directly on the sensory surface where the receptors are actually located[35]," whereas in every other system sensation is projected onto the external world. In fact a subset of the somatosensory receptors correspond to the internal viscera, and while this system still comprises an external sense in both Descartes' seven sense system and our own, it does begin to bridge the gap between internal and external sensation and is in many ways more intimate than the others. Anomalies in this system that can be understood in terms of brain function provide additional evidence to doubt the need of an extra-corporeal soul.

In a condition known as phantom limb syndrome, named and first clinically described by Silas Weir Mitchell, patients "experience an amputated extremity as still present, and in some cases also experience pain or cramping in the missing limb[35]." Somatosensory sensation is still perceived even in the absence of the physical limb, though "the patient recognizes that the sensations are not veridical, i.e., what he/she experiences is an illusion[35]." Interestingly, while we offer this condition and its physiological explanation as reason to discount the hypothesis of a soul, it has actually been employed historically to suggest the exact opposite interpretation:

> "After Lord Nelson lost his right arm during an unsuccessful attack on Santa Cruz de Tenerife, he experienced compelling phantom limb pains, including the sensation of fingers digging into his phantom palm. The emergence of these ghostly sensations led the sea lord to proclaim that his phantom was 'direct proof of the existence of the soul.' If an arm can survive physical annihilation, why not the entire person[35]?"

The Admiral was not alone in experiencing this bizarre phenomenon; in fact it occurs in ninety to ninety-eight percent of amputees following surgery, though in most cases the phantom sensations gradually fade after a few weeks or days.

In those cases where the phantom limb does persist, clinical observations range over a wide spectrum. Ramachandran notes that "phantoms are more vivid, and persist longer, after traumatic limb loss, or following amputation for a pre-existing painful limb pathology, than after a planned surgical amputation of a non-painful limb[35]." Patients often experience the sensation of pain in their phantom limbs. Perception of motor control ranges from paralysis; to a habitual posture, one patient who had a grenade explode in his hand retained this clenched and painful posture; to the ability to voluntarily move the phantom limb; to the perception of involuntary movement of the phantom. Moreover, "patients sometimes continue to feel a wedding ring or watch band on the phantom[35]." In one case, a patient "reported that, before amputation, the arthritic joint pains in her fingers would often flare up when the weather was damp and cold. Remarkably, whenever the air became humid the same pains would recur in her phantom fingers[35]."

In the absence of the limb, rearrangement occurs in somatosensory cortex and afferent axons from other body areas develop connections to cortical neurons that previously received input from the now missing limb. Just as there is a retinotopic map of the visual world, primary somatosensory cortex contains a somatotopic map, a representation of the touch receptors throughout the body laid out on this region of the brain. There is some distortion in this map; more processing power is devoted to regions with more sensitivity and areas that are adjacent on the body proper are not necessarily so in their cortical representation. For instance, in the somatotopic map, the upper limbs are adjacent to the face. The resulting reorganization after upper limb amputation documented by Ramachandran is exactly what one would expect from this type of organization; there are two sites that when stimulated induce sensation in the phantom limb corresponding to the regions on both sides of the now deafferented cortex. Patients report sensation in the phantom limb in response to stimulation of either the lower ipsilateral face or the region just above the amputation line. Moreover these areas of

referred sensations preserved a continuous and stable somatotopy. Stimulation near the amputation line was often more robust; Ramachandran describes a case where a patient would report sensation in the phantom limb when *he* stroked her face but not when she performed the stroking herself[35]. This suggests some form of inhibition perhaps because the "tactile sensations are perfectly coordinated with the patient's motor commands[35]."

Ramachandran constructed an experimental device called a *virtual-reality box* or a *mirror box* where a subject's intact limb is reflected so as to give the visual illusion of the non-existent contralateral limb. This has revealed some of the most interesting aspects of phantom limb syndrome. A patient, D.S., was able to overcome paralysis in a phantom arm by attempting to perform bilateral symmetric movements in the device; apparently the illusory visual feedback from the reflection of the existing arm was sufficient to liberate the phantom from its paralysis. When the mirror was removed, he reported the arm as once again "frozen as in a cement block[35]." This was repeated in a total of ten patients with results along a spectrum; some required maintenance of the visual illusion for motor control, some were able to learn to move the phantom limb with the subsequent absence of the device. Three observations involving the device prove especially illustrative. In one case, the phantom limb simply disappeared completely after repeated practice. In another, the device generated illusions whereby it appeared that a phantom finger was hyperextended into an anatomically impossible position, and the patient reported, with great surprise, actually experiencing this position. Finally, Ramachandran tested a finger amputee who had briefly experienced a phantom immediately after the amputation, but not in 28 years; the phantom returned instantly when the patient looked in the mirror device[35].

Other somatosensory illusions in non-phantom limb patients provide an interesting complement. In one, called the *phantom nose*, a subject is blindfolded and the experimenter uses the subject's hand to stroke the nose of a second subject sitting in front of him, while simultaneously using his own hand to stroke the initial subject's nose in perfect synchrony. The subject develops the powerful illusion that his own nose has either been dislocated or stretched several feet in front of him[35]. In another illusion, a subject is made to "project tactile sensations onto an inanimate object[35]," i.e. a shoe. The subject's hand is placed out of site under a table, and the experimenter strokes or taps the

subjects hand and a shoe placed on the table, again in perfect synchrony, and "after 10-30 seconds, the subject starts developing the uncanny illusion that the sensations are now coming from the shoe and that the shoe is now part of his body[35]." A powerful confirmation of the firmness of this illusion is that when the experimenter hit the shoe with a giant rubber hammer, the subject winced and registered an increase in skin conductance on a galvanic skin response test. Integrating these experiments with the observations of phantom limb syndrome suggests "that the so-called body image, despite all its appearance of durability and permanence, is an entirely transitory internal construct that can be profoundly altered by the stimulus contingencies and correlations that one encounters[35]."

The intricacies of phantom limbs and their relationship to body image suggest a conclusion quite opposite Lord Nelson's. Rather than simply the spontaneous firing of stump neuromas, as was originally thought, perception of a phantom limb actually corresponds to the result of integrating discrepant signals at quite a high level. "To generate any stable percepts (such as one's body image) or even a stable belief system the brain must weigh evidence from many different sources and quickly arrive at a decision[35]." This function, of "imposing coherence on information from diverse sources and for vetoing discrepancies," is performed by parietal cortex, in order "to confer stability on behaviour, avoid indecisive vacillation and optimize the allocation of one's cognitive and physical resources, given the ever-present need for rapid, effective action[35]." This explanation accounts for the clinical manifestations of phantom limb syndrome, incorporates a plausible evolutionary explanation, and illuminates a high-level processing system that performs some of the most complex functions of the human mind yet can be explained entirely by the brain.

In phantom limb-syndrome one perceives sensations that are not really there. The opposite of this, a tactile analog of blindsight[41], provides an interesting complement. Paillard, Michel, and Stelmach, a team of French neurologists, described a patient with a left parietal lobe defect resulting in tactile anesthesia on her right side. In perfect analogy to blindsight, she did not consciously perceive when the experimenters touched her right arm. This tactile irresponsiveness was complete; she did not respond "even to the strongest pressure[41]." However, when blindfolded and stimulated on her unfeeling right

arm and then asked to identify where she was touched, she was able to point to the location. Her response, translated from the French, was "I don't understand that! You put something here. I don't feel anything and yet I go there with my finger… How does that happen[41]?" This bears a resemblance to D.B.'s comments, though the somatosensory version of residual sensory capacity without conscious awareness is perhaps more elusive than its visual analog because of the nature of this sensory system.

The somatosensory system also carries information about pain, the perception of which is another intimate faculty. Pain perception is carried along a distinct pathway than the other somatosensory receptors, which can readily be demonstrated by hemisection of the spinal cord. A lesion to the right side will result in loss of pain perception on the left and loss of touch and proprioception on the right because the pain fibers cross immediately whereas the others do not. Pain signals are carried from receptors called nociceptors, which respond to a variety of painful stimuli, but pain and nociception are by no means functionally synonymous. The latter refers to the stimulation of the pain receptor and the former to the perception of receptor activation as pain, perhaps $pain_2$. This is one of the most dramatic loci in the nervous system of a dissonance between receptor activation and perception. Daniel Dennett argues that there is no such thing as $pain_1$, he claims that "no identification of pains or painful sensations with brain processes makes sense" and that "pains or painful sensations are 'things' discriminated *by people*, not for example, by brains[14]." The argument is interesting because it highlights the uniquely subjective nature of pain perception, but one must be careful with Dennett, and the author addresses his latter statement by maintaining that the 'brain' is *necessarily* the means by which 'people' 'discriminate pains.'

Pain is inherently subjective; it is often identified only by the subject's reporting feeling pain. The immediate counterarguments are that one can objectively identify nociceptor firing via neurophysiology, and that reflexive movements like withdrawal from painful stimuli can serve as indicators of pain perception in analogy to eye movements with vision. However we have already noted the significant dissociation between pain and nociception, and as Dennett rightly points out, a reflexive explanation is incomplete, "although simple withdrawal may be the basic or central response to such stimulation, in man and higher animals it is not the only one[14]." Different people have

different thresholds for pain and people may describe pain differently based on a lifetime of experience. A great degree of modulation occurs on the transduction of pain signals themselves, related to state of arousal, release of peptide neurotransmitters such as substance P that mediate hyperalgesia, and descending regulation whereby brain structures suppress nociceptor signaling. Finally there is the observation of referred pain, such as chest pain felt during myocardial ischemia, resulting from converging inputs from the viscera and the skin nociceptors at the spinal cord. These phenomena taken in concert perhaps make discussion of pain perception more elusive, but they do not render it inexplicable by the brain as Dennett suggests.

The Frontal Lobes:

By this point, localization of brain function, not absolute in the phrenologists' sense, but the implication of particular brain regions in particular functions, should have emerged as a motif. The occipital lobes are involved in visual processing, the temporal lobes in auditory processing, and the parietal lobes in somatosensory awareness. More specifically, a particular domain of this last region, the posterior parietal cortex, is involved in integrating multimodal information to generate coherent percepts and a sense of where the body is in space. The high-level function of posterior parietal cortex, already explored in the context of phantom limbs, is rivaled by the frontal lobes, which are responsible for a class of capacities collectively called executive function, including working memory, planning, decision-making, attention, and monitoring and updating one's own behavior[32, 33], which makes them a perfect place in the brain to look for Descartes' soul. The frontal lobes are extensively developed in humans, comprising as much as a third of the entire cortex[33], are much less so in other primates, and hardly at all in other mammals[37]. Moreover they continue to develop well after birth, physically growing until the age of 7[37] and refining their connections until the early 20's[33]. The frontal lobes make connections with nearly all other parts of the brain, including internal reciprocal connections, reciprocal connections with sensory cortex, and connections to the hippocampus, hypothalamus, amygdala, and anterior cingulate cortex[*]. These

[*] The hippocampus is involved in memory, the hypothalamus in regulation of homeostasis and neuroendocrine interaction, the amygdala in emotion, particularly fear response, and the anterior

observations fit their role as being "concerned not with the 'lower' functions of movement and sensation, but the highest ones of integrating all judgment and behavior, all imagination and emotion, into that unique identity that we like to speak of as 'personality' or 'self'[37]."

The same properties that make the frontal lobes likely candidates for many of the functions that Descartes attributes to the soul also render them less amenable to the methods of study that yielded much of the sensory information presented thus far. Having emerged so late on an evolutionary time scale, the frontal lobes are one of the structures where differences between humans and even our closest relatives may be most significant. A great deal has been learned about executive function from primate studies, but this mode, with respect to the frontal lobes, is inherently harder both because of the nature of the functions being investigated and the extent to which the animal model is valid. Likewise there are ethical limitations on what can be performed in humans. A further complication is that the functions attributed to the frontal lobes are less concrete than those of other cortical areas; the statement *V4 is involved in perception of color* is less abstract than *ventrolateral prefrontal cortex is involved in rule learning*[32]. It is also likely that many of these processes are more diffuse and involve numerous connections between various brain structures, and thus studying these functions by focal lesions, even if possible and ethically permissible, likely would still be less informative.

Frontal lobe lesions in humans do occur, and are the basis for much of the information that has been learned about their function. Brain damage is inherently messy and does not respect cytoarchitectonic boundaries; a railroad spike selectively destroying only Brodmann's area 46, dorsolateral prefrontal cortex, has about the same likelihood as an asteroid colliding with the Earth and entirely destroying the state of Massachusetts while leaving Connecticut, Rhode Island, Vermont, New Hampshire, and New York entirely unscathed. Moreover there are a relatively small number of cases, that are necessarily heterogeneous, and thus a meta-analysis is not useful. Instead, the historical strategy has been to look at these individual case studies and assess their similarities and differences. This is our strategy here, and the intention is that the idiosyncrasies and the

---

cingulate is discussed below, and is perhaps even more elusive than the frontal lobes themselves[33].

commonalities of this handful of patients will paint a picture of the frontal lobes such that we can ascribe to them many of the functions of Descartes' soul. Moreover the development of more sophisticated and higher resolution brain imaging technologies has enabled our knowledge of frontal lobe function to blossom tremendously in recent years. The survey begins with perhaps the most famous example of frontal lobe damage, Phineas Gage.

Phineas Gage:

Prior to the accident that radically transformed his personality, Phineas P. Gage, a twenty-five year old construction foreman for Rutland & Burlington Railroad, "had a well-balanced mind and was looked upon by those who knew him as a shrewd, smart businessman, very energetic and persistent in executing all his plans of action[12]," described by his bosses as the "most efficient and capable[12]" man in their employ. That would all change in the summer of 1848, when Gage was performing the routine task of blasting rocks in order to clear a path for the railroad through Cavendish, Vermont. In a freak accident where a member of Gage's crew failed to properly cover the charge with sand, rather than blasting the rock as was intended, a three foot seven inch, thirteen and a quarter pound iron bar projected backward and "enter[ed] Gage's left cheek, pierce[d] the base of the skull, traverse[d] the front of his brain, and exit[ed] at high speed through the top of his head[12]." Amazingly, the initial insult was minimal. Gage was awake, shortly able to speak, and capable of getting up and walking to the ox cart that carried him away. He was treated shortly after physician John Harlow, who would describe the case in his 1848 paper "*Passage of an iron rod through the head*[26]," and pronounce Gage cured after two months.

In the paper he would write twenty years later however, "*Recovery from the passage of an iron bar through the head*[27]," Harlow would reach a drastically different conclusion as the effects of Gage's injury on his personality manifested. He described the new Gage as

> "fitful, irreverent, indulging at times in the grossest profanity which was not previously his custom, manifesting but little deference for his fellows, impatient of restraint or advice when it conflicts with his desires, at times pertinaciously

> obstinate, yet capricious and vacillating, devising many plans of future operation, which are no sooner arranged than they are abandoned… A child in his intellectual capacity and manifestations, he has the animal passions of a strong man[27],"

in stark contrast to the 'shrewd, smart businessman' with a 'well-balanced mind,' "temperate habits and considerable energy of character[12]" whom he had been prior to the accident. More concise and perhaps more poignant is the statement given by an acquaintance that "Gage was no longer Gage[12]." After the accident, Gage was no longer able to secure a job; his former employers at Rutland & Burlington "considered the change in his mind so marked that they could not give him his place again" despite his reputation as one of the 'most efficient and capable' foremen. He met similar difficulty elsewhere, entertained a brief stint in the circus, and then returned to his mother and sister in California where he would die in 1861 in relative obscurity after experiencing a convulsion and never regaining consciousness.

Let us consider exactly what capacities or functions were disturbed in Gage's case. Harlow describes the post-accident Gage as 'fitful, irreverent,' and profane, which we can call social disinhibition, or failure to adhere to social rules and conventions that he had previously learned and followed. He was 'impatient of restraint or advice' and 'capricious and vacillating;' in other words, Gage demonstrated an inability to properly attend to something. Finally, while Gage apparently *made* plans for the future, they were 'no sooner arranged than… …abandoned.' Performing goal-directed behavior requires one first establish and identify a goal, formulate a plan to achieve the goal, hold the plan in working memory, and receive, assess, and respond to feedback in order to revise the plan if necessary. Planning for the future involves anticipating consequences, as does adhering to social conventions. In Gage this set of faculties was disturbed, and thus these abilities disappeared; "the equilibrium, or balance, so to speak, between his intellectual faculty and animal propensities had been destroyed[12]" and the new Gage, while perhaps intellectually intact, could not use these abilities to make appropriate social decisions.

Gage's injury did not simply result in a depression of function, but a fundamental shift, i.e., not from *A* to *a*, but from *A* to *B*. The analogy would be to a V4 lesion not resulting in loss of color vision, but a radically different experience of color vision, such

as suddenly seeing the world in fluorescence. One could conceive a situation where the same capacities were destroyed, and the resulting Gage was unable to empathize or plan for the future, but largely frozen, apathetic, and passive. That this was not the case reveals the intricacies of the frontal lobes. Rather than a subdued version of his old self, he acquired a fundamentally different personality; 'Gage was no longer Gage.' His decisions were "not the reserved or slight decisions of someone whose mind is diminished and who is afraid to act[12]," but rather equally assertive, just seemingly no longer guided by any sort of value system. Gage was not unable to make decisions, he was unable to make *good* decisions.

The most remarkable feature of Gage's case is the almost complete dissociation of the disturbance of personality and executive function from any other cognitive capacity. With the exception of blindness in his left eye, Gage had no sensory or motor deficits, and his perception, language, memory, and intelligence remained uncompromised. In retrospect, the pairing of a focal lesion and highly specific cognitive dissociation profile is strong evidence for localization of brain function, another instance of the motif that has guided our exploration of sensation. Mayr's warning about awareness of one's framework of ideas becomes relevant here, and in the nineteenth century, the scientific community was reluctant to accept this interpretation of Gage's case:

> "to understand Gage's behavioral change would have meant believing that normal social conduct required a particular corresponding brain region, and this concept was far more unthinkable than its equivalent for movement, the senses, or even language[12]"

However the inevitable conclusion was that "the observance of previously acquired social convention and ethical rules could be lost as a result of brain damage, even when neither basic intellect nor language seemed compromised[12]." Further,

> "Gage's example indicated that something in the brain was concerned specifically with unique human properties, among them the ability to anticipate the future and plan accordingly within a complex social environment; the sense of responsibility toward the self and others; and the ability to orchestrate one's survival deliberately, at the command of one's free will[12]"

The evidence seems to suggest that the capacities destroyed by Gage's injury, a set curiously similar to the functions Descartes reserved for the soul – social behavior, ethical judgment, planning, and perhaps even the will itself, were located in, highly dependent on, or at least related to the frontal lobes.

We now turn from Gage's cognitive profile to the physiological one. This picture is slightly complicated considering there was no autopsy performed; in fact, Harlow did not even learn of Gage's death for almost five years. However he did eventually recover Gage's skull, which, along with the iron now resides at the Warren Medical Museum at Harvard Medical School. Thanks to a little detective work on the part of Hanna Damasio, which involved photographing Gage's skull, using these measurements to construct a three-dimensional image of his brain, and then simulating the trajectory of the iron rod's path through his brain, we now have an understanding of exactly which brain regions were damaged. The nature of a penetrating injury is that it is focused and often the only damage is at the site where the tissue is actually destroyed, as opposed to the more widespread damage that would result from dissipation of force for a non-penetrating injury. This also explains way Gage remained conscious. Damasio's work reveals that Gage sustained damage to the ventromedial prefrontal cortex, and that the damage was more extensive on the left than the right side. The ventromedial prefrontal cortex (VM-PFC) has since been implicated in precisely these functions in humans and macaques. Its function is essential for normal performance in delay tasks, where subjects are required to forego an immediate reward for a larger reward later; decision making assessments such as the Iowa Gambling Task[†]; and conflict tasks where a subject must inhibit a prepotent response, for instance reading, in favor of a novel response, color identification, such as saying 'green' rather than 'red' in response to the word 'red' printed in green[32]. Antonio Damasio notes that VM-PFC damage "is accompanied by some disturbance of emotional behavior that includes both a diminished resonance of emotional reactions in general *and a specific compromise of social emotions such as*

[†] In this task, subjects are presented with four decks from which to choose cards. Every card carries some amount of reward, play money, but some also carry penalties. The object of the game is to win as much money as possible. Some of the decks are 'bad' decks and over the long run are associated with overall loss, whereas others are 'good' decks are associated with net gains. Normal subjects will after several trials identify this and only choose cards from the 'good' decks.

compassion and embarrassment[13]," whereas lesions of the adjacent dorsolateral prefrontal cortex (DM-PFC) "cause defects of decision but are not associated with the same emotional profile[13]," suggesting a highly specific role for this structure with regard to the intersection between social, emotional, and decision-making capabilities.

We now come nearly full circle and consider Gage's case from a philosophical perspective, reintroducing the concept of the soul, not the disembodied soul of Descartes, but rather as an entirely hypothetically and theoretical entity whose definition we postpone. Damasio questions "Gage's status as a human being[12]" after his accident. Indeed, 'Gage was no longer Gage,' but no longer human? He "lost something uniquely human, the ability to plan his future as a social being[12]." William Calvin notes the absence of such long-term intentional planning in all other animal species[8], and we add this to the candidate list of uniquely human capacities. However even if a given capacity X is exclusively human, we propose a tentative list including language, ethics, free will, and ability to plan for the future, and we strip these capacities away from the person one at a time, it seems silly to say there is a point at which they cease to be human. That is akin to saying that if you destroy a chameleon's ability to blend in with its surroundings it suddenly ceases to be a chameleon. However that does not preclude us from asking these questions, "[h]ow aware was [Gage] of this loss? Might he be described as self-conscious in the same sense that you and I are? Is it fair to say that his soul was diminished, or that he had lost his soul[12]?" Further, "May he be described as having free will? Did he have a sense of right and wrong, or was he the victim of his new brain design, such that his decisions were imposed upon him and inevitable? Was he responsible for his acts[12]?"

The author does not profess to have satisfactory answers for these questions. However the fact that we can pose them suggests that it is possible that these capacities can reside entirely in the brain and be explained entirely in terms of neurons and neural connections. Damasio asks a similar question to the one we posed earlier, "What would Descartes have thought had he known about Gage and had he had the knowledge of neurobiology we now have[12]?" Temporally, Gage's case is only the halfway marker between Descartes' time and our own, and already it seems likely that Descartes would no longer have a need to attribute the capacities of the *soul* to an extra-body entity. We

conclude our discussion of Phineas Gage by enumerating the faculties that were destroyed by his injury, and will return to this list to evaluate its degree of overlap with the functions of the soul after considering a few other cases. Damage to Gage's prefrontal cortex resulted in his inability to observe social conventions, behave ethically, make decisions advantageous to his survival, plan for the future, and perhaps to be aware of this deficit or exercise free will.

The Phineas Gage Matrix:

Damasio refers to this collection of frontal lobe patients with a similar cognitive profile – dissociated damage to social, emotional, ethical, volitional, and decision-making capacities as the 'Phineas Gage matrix'. He notes that "it is in the inevitable nature of syndromes to have a matrix, a shared essence of symptoms, and to have symptom variance around the edges of that essence[12]," hence the strategy here of surveying a few additional cases before making some generalizations about frontal lobe function and its relation to the alleged soul. In a case that preceded Gage, Dr. North, a master of Trinity College, Cambridge, was described as anxious and obsessive prior to a right hemisphere stroke, after which

> "His fears had left him. His scrupulosity, his diffidence, his seriousness, even his morality – all had vanished. He lay on his bed, in reckless levity, pouring forth a stream of flippant observations, and naughty stories, and improper jokes. While his friends hardly knew which way to look, he laughed consumedly, his paralyzed features drawn up in a curiously distorted grin[40]."

This case bears some resemblance to Gage; David Ferrier observed a somewhat different profile in removing the frontal lobes from monkeys:

> "Notwithstanding this apparent absence of physiological symptoms, I could perceive a very decided alteration in the animal's character and behaviour… Instead of, as before, being actively interested in their surroundings, and curiously prying into all that came within the field of their observation, they remained apathetic, or dull, or dozed off to sleep, responding only to the sensations or impressions of the moment[37]."

On the one hand, this is seemingly the exact opposite of Gage and Dr. North; the lobotomized monkeys were characterized by passivity and apathy, whereas the human patients experienced a qualitative personality shift. However Ferrier removed the entire frontal lobes, whereas, at least in Gage's case the damage was known to be more focal; perhaps selective disruption of more subtle faculties was masked by the larger and more global effect of completely removing the frontal lobes. Another possible implication is that over evolution the human frontal lobes have assumed a more specific function than even in our closest relatives. Perhaps interest in one's surroundings, selective attention and working memory[‡], and short-term planning capabilities comprise the totality of the monkey's executive function and as was the case with blindsight, an animal model can only take us so far in understanding a particular phenomenon.

There are also two developmental cases worth considering, one described by Donald Hebb and Wilder Penfield in 1940 and the other by S.S. Ackerly and A.L. Benton in 1948. The Hebb-Penfield patient suffered bilateral frontal lobe destruction after an accident at sixteen years of age, and his social development was arrested, and previously acquired social behavior deteriorated. The Ackerly-Benton patient sustained the injury shortly after birth, and despite normal cognitive development, his social development was abnormal. While generally amiable and polite, he was prone to outbursts and sudden loss of interest and indifferent to reward or punishment. Damasio notes the commonalities between these two developmental cases,

> "Rigid and perseverant in their approach to life, they both were unable to organize future activity and hold gainful employment; they lacked originality and creativity; they tended to boast and present a favorable view of themselves; they displayed generally correct but stereotyped manners; they were less able than others to experience pleasure and react to pain; they had diminished sexual and exploratory drives; and they demonstrated a lack of motor, sensory, or communication defects, and an overall intelligence within expectations[12]"

Note the inability to plan for the future or function socially, as with Gage; the diminished drives, as in Ferrier's monkeys and perhaps a function of the larger area damaged or the earlier onset; and the dissociation, which by now has become a nearly universal theme.

[‡] Which, in humans, is associated with the DL-PFC, which would have also been removed in Ferrier's experiments

Moreover, consider the questions that we asked of Gage and the extent of his humanity after the accident. They become even more difficult to answer in the context of the Ackerly-Benton patient, who never had a chance to develop these faculties. To round out our picture of the Phineas Gage matrix, we now turn to two additional case studies, the first by Antonio Damasio of a patient named Elliot and the second by Oliver Sacks of a patient named Greg F.

Elliot:

Damasio began seeing Elliot when he was in his early thirties. He had developed severe and debilitating headaches, and was eventually diagnosed with a midline meningioma, a tumor the size of an orange that compressed his frontal lobes upward[12]. The tumor was surgically removed with the concomitant removal of frontal lobe tissue. As was the case with Gage, the primary operation was a success and Elliot was *cured*; the collateral damage, the fundamental personality change, began during his physical recovery.

Elliot became less effective at his job and seemed to lack motivation and time-management skills. While the finer skills required for his job were spared, he would often become consumed with the minutia of a particular task, and was unable to prioritize or see the big picture; "he might spend a whole afternoon deliberating on which principle of categorization should be applied… date, size of document, pertinence to the case, or another[12]." With respect to work, "his actions were unnecessarily detailed… …at the expense of the overall purpose[12]." This deficit quickly translated into his personal life, and after losing his job Elliot began a sporadic series of business ventures, eventually resulting in bankruptcy, divorce, and a "brief marriage to a woman of whom neither family nor friends approved[12]" and a subsequent second divorce; in short, a series of bad decisions one after another, and of quite large magnitude and far-reaching consequences. Damasio comments that "the machinery for his decision making was so flawed that he could no longer be an effective social being[12]."

As with Gage, the profound dissociation is what makes cases like this truly fascinating. Damasio describes Elliot as

> "pleasant and intriguing, thoroughly charming but emotionally contained. He had a respectful, diplomatic composure, belied by an ironic smile implying superior wisdom and a faint condescension with the follies of the world. He was cool, detached, unperturbed even by potentially embarrassing discussion of personal events[12]."

This provides a sharp contrast to Gage, who was anything but 'pleasant' and 'emotionally contained.' Before turning to these differences though, consider those faculties of Elliot's that remained intact. His intelligence, awareness, and memory were unchanged. He exhibited normal, even superior performance on an IQ test and the Wechsler Adult Intelligence Scale. Damasio performed a series of tests specifically designed to assess executive function. Elliot had normal attention, a normal digit span, and normal working memory; Damasio notes that most patients with frontal lobe damage have abnormal performance on most, if not all of the tests that he performed on Elliot. Even on tests having to do with ethics, social convention, and moral value, Elliot's performance was normal.

What Damasio finally identified as Elliot's hallmark deficit was nearly total emotional detachment. He notes, "I found myself suffering more when listening to Elliot's stories than Elliot himself seemed to be suffering[12]." The intellectual aspect of Elliot's decision-making seemed unimpaired; the problem was rather in the execution, which Damasio attributes to this disaffectation.

> "Elliot was far more mellow in his emotional display now than he had been before his illness. He seemed to approach life on the same neutral note. I never saw a tinge of emotion in my many hours of conversation with him: no sadness, no impatience, no frustration with my incessant and repetitious questioning[12]."

To test this, he showed Elliot a series of images intended to elicit an emotional response, to which Elliot "could sense how topics that once had evoked a strong emotion no longer caused any reaction, positive or negative[12]." Aside from the primary observation of the disappearance of the capacity to respond emotionally, this highlights the extent of Elliot's dissociation – he was completely aware of this. Damasio challenges,

> "Try to imagine not feeling pleasure when you contemplate a painting you love or hear a favorite piece of music. Try to imagine yourself forever robbed of that

> possibility and yet aware of the intellectual contents of the visual or musical stimulus, and also aware that once it did give you pleasure. We might describe Elliot's predicament as *to know but not to feel*[12]."

Like Gage, Elliot's decision making was compromised. In contrast though, this capacity was perhaps even more dissociated, and his case suggests the importance of emotional affectation on effective decision making. Elliot's damage was also to the VM-PFC, though his was more extensive on the right side, perhaps suggesting some lateralization of function in prefrontal cortex. We now add the experience of emotion and its apparent interconnectedness to rational decision making to frontal lobe functions that can be selectively disturbed. Moreover with regard to Elliot the question of free will is settled somewhat less ambiguously. The dissonance between his intellectual decision-making and his performance on Damasio's tests, compared to his real-world decision making and actual execution, is apparent, and one could feasibly conclude, which Damasio does, that Elliot lacked free will. This allows us to perform an interesting exercise. Consider the statement *emotions have the capacity to affect the will*, which Elliot's example readily demonstrates. Substituting some Cartesian phrases, 'passions' for 'emotions' and 'soul' for 'will' merely to acknowledge that for Descartes the former contains the latter, gives us *passions have the capacity to affect the soul*. Not remarkable in and of itself, but this is essentially the thesis of Descartes' *Passions of the Soul*, perhaps the most esoteric of all of his writing, and here we explain it entirely in terms of the brain.

Greg F.:

Elliot's case provided an example of an even more focal disturbance than Gage and allowed us to precisely consider the dissociation of emotional affect and rational decision making. We now consider quite the opposite situation, with far more diffuse damage and a broader spectrum of impairments, a patient by the name of Greg F., and the subject of Oliver Sacks' *The Last Hippie*[37].

Like Elliot, Greg suffered from a midline tumor, though the diagnosis of his was impeded by his staying at a Hare Krishna temple where his diminishing vision, physical changes, disorientation, and strange smile were interpreted as 'spiritual progress.' By the

time it was diagnosed, Greg's tumor was the size of a grapefruit, and had destroyed his pituitary gland, optic chiasm, optic tract, and major parts of his frontal and temporal lobes. Destruction of his pituitary gland disrupted basic regulation of homeostasis; Greg became "fat and hairless." From the destruction of the optic tract, Greg was totally blind. He exhibited severe anterograde and retrograde amnesia. His tumor was removed in 1976, and he had no memory of anything since 1970, and his memory from prior to that was graded and improved as one went further back in time; memories from 1966 were largely intact. His anterograde amnesia was even more pronounced; he could not recall a list of words after about a minute. This is consistent with the destruction of Greg's hippocampus, involved in forming new declarative memories, though Greg was able to slowly habituate, to which Sacks implicates a more primitive memory system. An interesting feature of Greg's amnesia was the singular exception of music; Greg was able to learn new songs and pick up limericks and jingles with ease.

With brain damage as extensive as Greg's we face the danger of larger disturbances masking the more subtle ones. Sacks notes that "widespread damage had created a very complicated clinical picture, with sometimes overlapping or even contradictory symptoms and syndromes[37]." Greg was described as "Unaware – and indifferent. He seemed bland, placid, emptied of all feeling[37]." While this resembles Ferrier's monkeys and Elliot, and could be attributed to frontal lobe damage, Sacks notes this could also be a manifestation of pituitary damage "undermining his hormonally driven aggressiveness and assertiveness[37]." He continues, describing Greg's state as idle, "almost devoid of mental content or affect," an "intermediate, half-dreamlike state in which, if the normal control and selectivity of thinking was lost, there was a half freedom, half compulsion, of fantasy and wit," with "elements of the primitive, the childlike, the playful;" an "appearance of innocence and wisdom combined[37]." Yet he became animated by social contact, and was "invariably cheerful, euphoric[37]."

Greg's blindness was accompanied by blindness denial; while he admitted his vision was not very good, he still professed he could see, though he was in fact completely blind. Sacks would show him an object, for instance, a green comb, and Greg would confidently identify it as a blue ball. He professed that he enjoyed watching television. As Greg progressed, they tried to teach him Braille, at which he became

frustrated and demanded, "What's going on? Do you think I'm blind? Why am I here, with blind people all around me? If I were blind, I would be the first person to know it[37]."

Moreover Greg exhibited something called *witzelsucht*, or joking disease, a characteristic symptom of orbitofrontal patients, and perhaps a milder manifestation of what was observed in Gage and Dr. North. It is characterized by reacting "immediately and incontinently to everything around them and everything within them – to virtually every object, every person, every sensation, every word, every thought, every emotion, every nuance and tone" with some loss of restraint, caution and inhibition and a propensity to "wordplay and puns[37]."

This is perhaps a particular manifestation of what the author considers the single defining feature of Greg's case – "his confinement, in effect, to a single moment – "the present" – uninformed by any sense of a past (of a future)[37]." Sacks identifies "something stranger, and more complex, than a mere "deficit," rather some radical alteration within him in the very structure of knowledge, in consciousness, in identity itself[37]." He continues,

> "Given this radical lack of connection and continuity in his inner life, I got the feeling, indeed, that he might not *have* an inner life to speak of, that he lacked the constant dialogue of past and present, of experience and meaning, which constitutes consciousness and inner life for the rest of us. He seemed to have no sense of "next" and to lack that eager and anxious tension of anticipation, of intention, that normally drives us through life[37]."

In comparison to Gage or to Elliot, Greg is perhaps the most debilitated, and if one were to arrange the elements of the Phineas Gage matrix along a spectrum and pose the set of questions that we did for Gage, Greg may be the easiest for whom to conclude that his humanity[§] has been lost. However Sacks' comment is especially pertinent, the author concludes this section with Sacks' words and postpones his own commentary until after the consideration of a final case:

> "Greg's 'frontal lobe' characteristics – his lightness, his quickfire associations – were fun, but beyond this there shone through a basic decency and sensitivity and

[§] Here referring to uniquely human faculties, not his status as a person.

> kindness. One felt that Greg, though damaged, sill had a personality, an identity, a *soul*[37]."

The Anterior Cingulate and the Will:

Antonio Damasio and Francis Crick both suggest that the anterior cingulate cortex is a candidate brain structure for the seat of the will. We briefly explore this hypothesis in the context of a patient of Damasio's, Mrs. T., not to support or reject it, but to assess to what extent we can ascribe a physiological basis to this last faculty, the will, which is highly intimate, perhaps uniquely human, and the one that Descartes most resolutely reserved for the disembodied soul.

Mrs. T. suffered a bilateral stroke that damaged regions of the anterior cingulate cortex, the supplementary motor area (SMA, or M2), and another motor area called M3[12]. After the stroke, Mrs. T became "motionless and speechless, and she would lie in her bed with her eyes open but with a blank facial expression[12]." She was attentive yet irresponsive, able to

> "follow people with her eyes but [she] did not speak spontaneously. She gave no verbal reply to any questions put to her even though it appeared she understood them because of the way she nodded in reply. She could repeat words and sentences but only very slowly[11]."

Eventually "she gradually emerged from this state of mutism and akinesia[12]," and it is her subsequent reflection on this period of 'suspended animation' that is most interesting.

> "Contrary to what one might have thought, her mind had not been imprisoned in the jail of her immobility. Instead it appeared that there had not been much mind at all, no real thinking or reasoning. The passivity in her face and body was the appropriate reflection of her lack of mental animation. At this later date she was certain about not having felt anguished by the absence of communication. Nothing had forced her not to speak her mind. Rather, as she recalled, "I really had nothing to say[12]."""

Contrast this with a patient with an active mind but motor, or even just verbal impairment who is merely unable to communicate. While the motor damage could have masked the more subtle underlying deficit, in Mrs. T.'s case it seems this was not the case. Crick claims, "Her mind had been 'empty,' she'd lost her Will[11]," Damasio that "her will had

been preempted… …there was a pervasive impairment of the drive with which mental images and movements can be generated[12].”

The anterior cingulate is involved in movement, emotion, and attentiveness; “damage to this sector causes a virtual suspension of the animation of action and of thought processes such that reason is no longer viable[12].” Lesioned patients showed decreased spontaneous behavior[2], producing fewer verbal utterances during interviews, shorter written statements, and fewer and simpler constructions when asked to play with Tinker Toys in comparison to controls. The anterior cingulate is also implicated in performance monitoring, a critical component of executive function whereby behavior is modified based on environmental feedback[32]. A study in monkeys reveals the role of this structure in mediating behavioral modification in response to changing reward contingencies. Monkeys were trained to perform two different tasks, either pushing or turning a handle, and were rewarded differentially for performing the tasks; when the reward decreased, the monkeys switched to the other task. Researchers identified a population of neurons in the anterior cingulate that responded to decreasing reward. When the monkeys failed to switch tasks, these neurons did not fire, suggesting a specific brain mechanism that identifies decreasing reward and initiates a change in behavior, essentially volition, or the will. In addition to the connections to prefrontal cortex and adjacent supplementary motor cortex, the anterior cingulate also connects extensively with the limbic system, leading Damasio to identify it as the intersection between emotion, reason, and external action[12]. Crick notes extensive reciprocal callosal[**] connections and that each anterior cingulate projects to motor areas in both hemispheres, employing arguments similar to those that Descartes did for the pineal gland to implicate this structure as the seat of the will[11].

The Gage Matrix Reloaded:

Having surveyed each of the elements of the so-called Phineas Gage matrix, we now consider the faculties associated with the frontal lobes. From Gage himself we have adherence to social convention, ethical behavior, making decisions advantageous to one’s

[**] Meaning that the anterior cingulate on one side projects across the corpus callosum, the large fiber tract connecting the two brain hemispheres, to the corresponding structure on the other side

well-being, planning for the future, and self-awareness. Dr. North's case underscores this first capacity, but we will call his contribution to the matrix social inhibition or restraint. The developmental cases of Hebb and Penfield and Ackerly and Benton add the ability to learn appropriate social behavior, experience of pleasure and pain, and sexual and exploratory drives. Elliot contributes the ability to experience emotion, which as his case reveals is critical to efficacious decision making. Greg F.'s profile of cognitive deficits is extensive, but his contribution to the Gage matrix is temporal awareness, not in the context of his amnesia, but his inability to detach himself from the immediate moment and to use knowledge of the past and future to influence his actions. Finally we have Mrs. T., whose case may get at the will in its raw form, i.e. the ability to initiate a volition, whether it manifest internally as a conscious thought or externally as a motor command[††]. A survey of frontal lobe cases reveals that each of these capacities can be perturbed solely by physical brain damage, and that therefore they can all be accounted for, though perhaps not yet adequately explained, by neurons and their interactions within the physical brain.

Williams Syndrome:

Finally we turn to Williams syndrome, a rare genetic disorder that provides an additional window into brain function. This disorder is characterized by a unique, dissociated, and highly stereotyped cognitive and social profile. While not an attempt to qualify the extent of genetic determinism, the discussion of Williams syndrome and the implication that a particular genotype is highly correlated with specific complex personality and social attributes, intends to reinforce the ongoing current with a genetic perspective, an essential element of post-Cartesian science.

Williams syndrome is a rare neurodevelopmental genetic disorder that corresponds to a hemizygous deletion of between twenty and twenty eight genes on the

[††] The author likes the analogy of Mrs. T's case as a more general application of the destruction of the basal ganglia in Parkinson's disease manifesting as difficulty *initiating* voluntary movements. Normally, a motor loop through this structure, the basal ganglia, is required to initiate movement – activation, or more accurately overcoming inhibition, must be achieved above a certain threshold for a motor command to be initiated. Consider the anterior cingulate as the analog of the basal ganglia and volition in its more general form as analogous to a motor command. Likely the analogy only holds superficially if at all but the author finds it useful in illustrating this point.

long arm of chromosome seven, chromosome band 7q11.23[7, 34]. While genetic, it is not heritable; the deleted region in Williams syndrome, which is the same in ninety-eight percent of the one in twenty-thousand people who suffer from the disorder[29], is flanked by regions of highly repeated DNA, and mispairing of these repeats during meiosis results in the deletion and resulting Williams syndrome phenotype. It is this homogeneity, both genotypic and phenotype and mechanistic elegance and simplicity that makes the disorder relevant to our present discussion. Julie Korenberg, one of the leading scientists presently studying Williams syndrome comments

> "here we have this really tiny genetic deletion — of the 20-some-odd genes missing, probably just 3 to 6 create the cognitive and social effects — that reliably creates a distinctive behavioral profile. Williams isn't just a fascinating mix of traits. It is the most compelling model available for studying the genetic bases of human behavior[21]."

A corollary is that this compelling model allows us to ascribe a genetic, and therefore physiological, basis to these aspects of behavior.

It is important to realize however the complexity of genetic regulation of behavior. It is not as if one of the twenty or so genes absent in Williams syndrome corresponds to *fear response to threatening faces*, and its absence results in the decreased amygdala activation observed in Williams patients in response to this stimuli compared to controls[34]. Rather some of the genes are patterning genes, which have a global effect during embryonic development and according to Albert Galaburda, this abnormal brain development ultimately results in an imbalance between dorsal and ventral regions and their corresponding functions. The former includes functions related to vision, spatial awareness, and intention, areas of noted weakness in Williams patients, whereas the latter includes language, auditory processing, facial recognition, emotion, music, and social drive, all of which are spared or even elevated[21]. The role of experience and environment cannot be understated however; these pattering genes set in motion a genetic program with slightly different constraints, and thus brain development, social behavior, and identity necessarily take a different course. Williams syndrome patients exhibit an 11% reduction in overall cerebral volume compared with controls, with selective losses in the thalamus, superior colliculus, parahippocampal gyrus, and occipital and parietal lobes[36].

On the other hand, they showed increases in gray matter volume and density in several brain regions, including the amygdala, orbital and medial prefrontal cortices, anterior cingulate, fusiform gyrus, insular cortex, and superior temporal gyrus. Ursula Bellugi also notes relatively normal cytoarchitecture, but decreased neuronal cell density, especially in the parietal regions[7]. What emerges from this picture is a disorder characterized by a specific genotype and a relatively homogenous alteration in brain structure.

The clinical diagnosis is equally stereotypical. Williams syndrome was first identified in 1961 by Dr. J. C. P. Williams of New Zealand who noted common symptoms in a series of cardiac patients. Cardiovascular abnormality is the most serious physiological symptom; deletion of the elastin gene results in supravalvular aortic stenosis (SVAS)[34]. Williams syndrome is also characterized by failure to thrive in infancy; transient neonatal hypercalcemia; delayed language and motor milestones, and abnormal sensitivities to classes of sounds, or hyperacusis[7]. People with Williams syndrome also exhibit a distinct craniofacial dismorphology characterized by a "broad brow, full cheeks, stellate iris, flat nasal bridge, full nasal tip, long filtrum, prominent lips and ear lobes, small, widely spaced teeth, and wide mouth[29]." They also have marked cognitive deficits, with IQ scores ranging from 40-100 with a mean around 60[7]. Consistent with the reduction of brain volume in particular regions, visuospatial impairment is the predominant cognitive deficit in Williams syndrome.

However the most interesting aspect of the disorder is the dissociation between spatial cognition, where there is profound impairment, and language and face processing, which are relative strengths, and the resulting behavioral profile termed the Williams social phenotype[7, 21]. Bellugi, one of the pioneers in describing the disorder, comments on her patients, "I didn't have to talk to them long to realize something special was going on. Here they had these great cognitive deficits. Yet they spoke with the most ardent and delightful animation and color[21]." Alan Reiss observes a similar profile of "disproportionately severe visual-spatial deficits and enhanced emotionality and face processing[36]."

The Williams social phenotype is characterized by "indiscriminate friendliness, enhanced emotional empathy, and loquaciousness among adults[29];" according to Bellugi,

"The language just seemed to be erupting out of them[21]." Or the account by a teenage girl with Williams syndrome, 'Everyone in the world is my friend.' Language use in those with Williams syndrome is complex. On the one hand language is clearly a *relative* strength; note the loquaciousness in adults and the richer vocabulary to age matched controls among children. However language development is not entirely normal; it actually begins later and likely follows a different trajectory than in normal subjects. Most notable are errors with spatial prepositions[7], 'near,' 'around,' 'in,' 'under,' and excessive use of evaluative language[29]. When asked to tell a story from a picture book, Williams subjects inserted far more examples of this evaluative language – enrichment devices, exclamatory phrases, and affective qualities than controls[7].

Also elusive is the relationship between language and sociability. An early hypothesis was that sociability drives language[21], though the relationship is likely to be more complex. The "love of company and conversation" characteristic of Williams syndrome is "combined, often awkwardly, with a poor understanding of social dynamics and a lack of social inhibition[21]." Williams subjects have a tremendous drive for social contact and interaction, but their social behavior is not always appropriate[7]. In addition to those features peculiar to language use already noted, the Williams social phenotype is characterized by difficulty comprehending affect; "lost on them are many meanings, machinations, ideas and intentions that most of us infer from facial expression, body language, context and stock phrasings[21]." Despite facial processing and recognition being a relative strength, they are unable to infer intention from facial expression. In fact, "cognitive scientists argue over whether people with Williams have theory of mind[21]." Also in contrast to the heightened social drive and overall gregariousness is simultaneous increase in social anxiety[36].

This anxiety however should not be confused with fear; in fact, the absence of social fear is another fundamental characteristic of the Williams social phenotype. Subjects displayed decreased amygdala activity compared to controls when presented with pictures of threatening faces, though there was no difference when both groups were presented with threatening objects[21]. This is supported anecdotally as well; Williams subjects are more inclined to approach strangers and parents of children with Williams syndrome reported this as well on a questionnaire administered by Bellugi et al.[29]. This

has been attributed to decreased amygdala activation by the orbitofrontal cortex in response to stimuli[21, 29]. The orbitofrontal cortex, part of the frontal lobes, "is associated with prioritizing behavior in social contexts[21]." Abnormal development here results in abnormal connections between this region and the amygdala, and thus "abnormal neural circuitry underlying social–emotional information processing in Williams syndrome[29]." Järvinen-Pasley et al., describe this lack of orbitofrontal activation and increased activity in medial prefrontal cortex and suggest its role in abnormal emotional behavior in Williams syndrome. Moreover she draws an explicit parallel of this deficit in this context to its manifestation with frontal lobe damage.

Williams syndrome exemplifies a specific genetic alteration that manifests as a specific behavioral phenotype. The interplay between genetic, epigenetic, developmental, environmental, and experiential factors is undoubtedly complex, but the relative homogeneity of the disorder at both ends – the underlying cause, a mere twenty-eight genes, and the stereotypical profile, provides an elegant example of genetic control over behavior. Moreover the behavioral alterations are subtle and complex, similar to those that we surveyed in discussing frontal lobe patients. In analogy with the focal lesions employed to glean further information in those cases, Järvinen-Pasley identified a case with an atypically small deletion where the patient had many of the characteristics of Williams syndrome but lacked the social phenotype, further reinforcing that it is genetically influenced and also honing in on the specific genes responsible. This disorder, in its elegance and simplicity, provides a "compelling model for elucidating the relationships between cognition, the brain and, ultimately, the genes[7]," and a genetic dimension to our quest to render Descartes' mind superfluous.

Conclusions:

The question that remains is whether the preceding examples have sufficient explanatory power to allow us to dispense with the hypothesis of a disembodied soul and to conclude that there is no "absolute or metaphysical limit on the possibility of scientific explanation[4]," but rather that each of our unique identities as an individual, consisting of a body and a mind, is entirely corporeal and that the latter can be explained entirely in terms of the brain.

The goal of this investigation has been to demonstrate that there actually exists an underlying harmony between Descartes and twenty-first century neuroscience. The first essential feature of this task is qualifying Descartes theory of mind and articulating it in the context of his entire ideology, which we have sought to do. The tenets that we have extracted from returning to the primary Cartesian texts include the primacy of establishing reliable knowledge, an affinity for true explanation rather than mere displacement, and most importantly, a trend toward reducing the domain of the soul rather than expanding it. Mind-body substance dualism is largely something that has been anachronistically read into Descartes, and while to some extent the *cogito* does necessitate that one look at the mind and body as distinct conceptual entities because of the inherent difference in how we can understand them, it does not necessitate that they are truly separate entities and that the former cannot be encapsulated by the latter. Viewing Descartes through the lens of substance dualism is in fact quite vehemently anti-Cartesian in its broadest sense; while Descartes arrived at his conclusions by 'razing everything to the ground' and erecting new foundations, this approach takes a particular implication of the theory and renders it a first principle.

Clark asserts that "the Cartesian theory of mind is a first rather bold step in the direction of removing souls completely from explanations of human behaviour[9]," and this author maintains that Descartes himself likely did not anticipate or intend this, but that with the understanding of evolution and cell theory and the explosion of knowledge in neuroscience, we can now take the next step in this process and completely reject the hypothesis of a disembodied soul. Francis Crick observes that "the history of science is littered with statements that something was inherently impossible to understand.[11]" However as Crick's own work in identifying the structure of DNA exemplifies, such statements are continually proven false and reflect an inadequate context for properly formulating a particular question rather than an inherent inaccessibility. In Descartes' time, a physiological understanding of the human mind was an enigma; in our own, it is more than feasible.

Discussion of theory of mind inevitably forces one to consider some of the philosophical questions that arise out of entertaining such an intimate capacity, and the implications of the answers to some of these questions have likely contributed to the

persistence of the disembodied soul hypothesis. For whatever reason, humans have an affinity for the inexplicable and mysterious, equating this unintelligibility with sublimity. One gets the sense that being able to explain the entirety of our mental life in terms of neurons somehow renders it less impressive and diminishes our humanity. This author fundamentally rejects that notion on the same grounds that understanding evolution as the mechanism of speciation renders it no less wonderful. In fact, the author asserts the contrary, that understanding is actually enriching and secondarily adds that such a fear is unnecessary, as we are a long way off from completely understanding the human mind.

The philosophical dilemma is to identify what makes us uniquely human and how. The Cartesian solution to this problem was to assign those capacities to the soul and assert that only humans have souls; the set includes some degree of conscious perception, executive function, and the will. Descartes asserts, "the control of animal inclination by thought, reason, and the will was what made us human;" Antonio Damasio's response to this underscores the crux of the present argument:

> "I agree with his formulation, except that where he specified a control achieved by a nonphysical agent I envision a biological operation structured within the human organism and not one bit less complex, admirable, or sublime[12]."

Damasio says elsewhere that

> "The fact that acting according to an ethical principle requires the participation of simple circuitry in the brain core does not cheapen the ethical principle. The edifice of ethics does not collapse, morality is not threatened, and in a normal individual the will remains the will[12]"

Once one can dispense with the prejudice that an understanding of a phenomenon renders it less spectacular, the task of reconciling seventeenth century philosophy with twenty-first century neuroscience and beginning to understand the physiological mind becomes less arduous. Rather than antagonistic, Cartesian philosophy and modern neuroscience are actually quite in accordance and ideas from the former can prove useful in the development of the latter. We can now move the problem of understanding the physiological mind from the third category of Descartes' *Regulae*, where the question

could not yet be formulated[*] to the second category, where the problem has not been solved but at least the question has been posed and we have opened the black box of the mind-brain duality that has been remained closed for four centuries and begun to investigate its mechanisms. This search will lead to further black boxes, but the fundamental difference is how we address the problem, rejecting the notion of an 'absolute or metaphysical limit on the possibility of scientific explanation' just as when Laplace described his account of the universe to Napoleon, and he inquired as to the role of God in Laplace's system, Laplace replied "'Je n'avais pas besoin de cette hypothèse-là," or "I have no need of that hypothesis."

[*] Ironically, Descartes never finished writing this third section.

Bibliography:

1. Aarsleff, Hans. 'Descartes and Augustine on Genesis, Language, and the Angels', in M. Dascal and E. Yakira (eds), *Leibniz and Adam*. Tel Aviv: University Publishing Projects, 1993
2. Allman, John M., et al. (2001) "The Anterior Cingulate Cortex: The Evolution of an Interface between Emotion and Cognition." Annals New York Academy of Sciences 935, 107-117
3. Almog, Joseph. What Am I? Descartes and the Mind-Body Problem. New York: Oxford University Press, 2002
4. Baker, Gordon and Morris, Katherine J.. Descartes' Dualism. London: Routledge, 1996
5. Barlow H B, 1972, "Single units and sensation: A neuron doctrine for perceptual psychology?" *Perception* **1**(4) 371 – 394
6. Behe, Michael J. Darwin's Black Box: The Biochemical Challenge to Evolution, 10th Anniversary Edition. New York: Free Press, 2006 (originally published 1996)
7. Bellugi, Ursula, et al. (1999) "Bridging cognition, the brain, and molecular genetics: evidence from Williams syndrome." Trends in Neurosciences, (22) 5, 197-207
8. Calvin, Bill. How Brains Think. New York: Basic Books, 1996
9. Clarke, Desmond M. Descartes's Theory of Mind. New York: Oxford University Press, 2003
10. Cottingham, John. editor. Reason, Will, and Sensation: Studies in Descartes's Metaphysics. Oxford: Clarendon Press, 1994
11. Crick, Francis. The Astonishing Hypothesis: The Scientific Search for the Soul. New York: Charles Scribner's Sons, 1994
12. Damasio, Antonio. Descartes' Error: Emotion, Reason, and the Human Brain. London: Penguin Books, 1994
13. Damasio, Antonio, 2007, "Neuroscience and Ethics: Intersections" *The American Journal of Bioethics* 7:1, 3-7
14. Dennett, Daniel C. Content and Consciousness. Boston: Routledge & Kegan Paul, 1969
15. Dennett, Daniel C. Darwin's Dangerous Idea: Evolution and the Meanings of Life. New York: Simon & Schuster, 1995
16. Descartes, Rene. Discourse on the Method of Rightly Conducting One's Reason and of Seeking Truth in the Sciences. Cress, Donald A. translator. Indianapolis: Hackett Publishing Company, 1980 (originally published 1637)
17. Descartes, Rene. Meditations on First Philosophy, Third Edition. Cress, Donald A. translator. Indianapolis: Hacket Publishing Company, 1993 (originally published 1641)
18. Descartes, Rene. Principles of Philosophy. Miller, Valentine Rodger and Miller, Reese P.. translators. Dordrecht, Holland: D. Reidel Publishing Company, 1983 (originally published 1644)
19. Descartes, Rene. The Passions of the Soul. Voss, Stephen. translator. Indianapolis: Hackett Publishing Company, 1989 (originally published 1649)

20. Descartes, Rene. The World and Other Writings. Gaukroger, Stephen. translator and editor. Cambridge University Press, 1998
21. Dobbs, David. "The Gregarious Brain". The New York Times Magazine. 8 July 2007
22. Dobzhansky, Theodosius. 'Nothing in Biology Makes Sense Except in the Light of Evolution.' American Biology Teacher, Volume 35 (1973) pp. 125-129
23. Gelder, Beatrice et al. Intact navigation skills after bilateral loss of striate cortex. Current Biology. 2008 Dec 23;18(24):R1128-9.
24. Gigerenzer, Gerd. Gut Feelings: The Intelligence of the Unconscious. London: Viking Penguin, 2007
25. Grierson, Bruce. U-Turn: What If You Woke Up One Morning and Realized You Were Living the Wrong Life? New York: Bloomsbury, 2007
26. Harlow, J. M. (1848). Passage of an iron rod through the head. *Boston Medical and Surgical Journal*, *39*, 389-393.
27. Harlow, J.M. (1868). Recovery from the passage of an iron bar through the head. *Publications of the Massachusetts Medical Society*, *2*, 327-347.
28. Hawking, Stephen. A Briefer History of Time. New York: Bantam Dell, 2005
29. Järvinen-Pasley, Anna, et al. (2008) "Defining the social phenotype in Williams syndrome: A model for linking gene, the brain, and behavior." Development and Psychopathology. 20, 1-35
30. Mayr, Ernst. The growth of biological thought. Cambridge: Harvard University Press, 1982
31. Ogren, Marilee. "Blindsight." Lecture, Neurobiology of Sensory and Motor Systems. Boston College. Chestnut Hill, MA. 24 February 2009.
32. Ogren, Marilee. "Executive Function." Lecture, Developmental Psychobiology. Boston College. Chestnut Hill, MA. 30 October 2008.
33. Ogren, Marilee. "Prefrontal Cortex." Lecture, Developmental Psychobiology. Boston College. Chestnut Hill, MA. 28 October 2008.
34. Ogren, Marilee. "Williams Syndrome." Lecture, Developmental Psychobiology. Boston College. Chestnut Hill, MA. 2 December 2008.
35. Ramachandran, V. S. and W. Hirstein. "The perception of phantom limbs. The D. O. Hebb lecture.." *Brain* 121 (Pt 9) (September 1998): 1603-1630.
36. Reiss, Alan L., et al. (2004) "An Experiment of Nature: Brain Anatomy Parallels Cognition and Behavior in Williams Syndrome." The Journal of Neuroscience. 24(21):5009-5015
37. Sacks, Oliver. An Anthropologist on Mars: Seven Paradoxical Tales. New York: Vintage Books, 1995
38. Sacks, Oliver. Musicophilia. New York: Vintage Books, 2007
39. Sepper, Dennis L. Descartes's Imagination: Proportion, Images, and the Activity of Thinking. Berkeley: University of California Press, 1996
40. Strachey, Lytton. "The Life, Illness, and Death of Dr. North." In *Portraits in Miniature*, 29-39. London: Chatto & Windus, 1931
41. Weiskrantz, L. Blindsight: A Case Study and Implications. Oxford: Oxford University Press, 1986