



2011

Religion, Science, and the Conscious Self: Bio-Psychological Explanation and the Debate Between Dualism and Naturalism

Paul J. Voelker
Loyola University Chicago

Recommended Citation

Voelker, Paul J., "Religion, Science, and the Conscious Self: Bio-Psychological Explanation and the Debate Between Dualism and Naturalism" (2011). *Dissertations*. Paper 242.
http://ecommons.luc.edu/luc_diss/242

This Dissertation is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Dissertations by an authorized administrator of Loyola eCommons. For more information, please contact ecommons@luc.edu.



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License](https://creativecommons.org/licenses/by-nc-nd/3.0/).
Copyright © 2011 Paul J. Voelker

LOYOLA UNIVERSITY CHICAGO

RELIGION, SCIENCE, AND THE CONSCIOUS SELF:
BIO-PSYCHOLOGICAL EXPLANATION AND THE DEBATE BETWEEN
DUALISM AND NATURALISM

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE GRADUATE SCHOOL
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

PROGRAM IN THEOLOGY

BY

PAUL J. VOELKER

CHICAGO, ILLINOIS

MAY 2011

Copyright by Paul J. Voelker, 2011
All rights reserved.

ACKNOWLEDGEMENTS

Many people helped to make this dissertation project a concrete reality. First, and foremost, I would like to thank my advisor, Dr. John McCarthy. John's breadth and depth of knowledge made for years of stimulating, challenging conversation, and I am grateful for his constant support. Dr. Michael Schuck also provided for much good conversation during my time at Loyola, and I am grateful to Mike for taking time from a hectic schedule to serve on my dissertation committee. Dr. J.D. Trout graciously agreed to serve on both my comprehensive exam committee and my dissertation committee, and I am the better for it. J.D.'s own research provided an excellent model of how to do interdisciplinary, naturalistic philosophy.

I am grateful to a number of my fellow grad students for help and support, especially Denise Starkey, Emily Reimer-Barry, and Ana Bedard. Catherine Wolf and Marianne Wolfe were a constant source of administrative help. I am grateful to the Graduate School at Loyola University Chicago for funding for an Advanced Dissertation Fellowship that enabled the completion of this project.

Friends and family were also a constant source of encouragement. In Chicago, many thanks to Adam Shames, Andy Weiss, and members of the Voelker and Fuerst clans, especially Heidi and Frank Hutton and Kate and Dave May. My late uncle and aunt, Werner and Marie Fuerst, were an unfailing source of hospitality during my time in Chicago; they are missed. I feel that my gang of friends from high school—the

Westsylvania— all contributed to this project in one way or another; special thanks go to Jeff Frank and John Christodouleas. Thanks also to Jim, Julie, and Anna Voelker.

Thanks in particular to Jim, my brother and actuary extraordinaire, for constant support and encouragement. Last but not least, thanks to my parents, Jacob and Rosemary Voelker, for being the first and best educators in my life. It is to them that this dissertation is dedicated.

For my parents, Jacob and Rosemary Voelker

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
ABSTRACT	ix
INTRODUCTION	1
CHAPTER ONE: NATURALISM, DUALISM, AND BIO-PSYCHOLOGICAL EXPLANATION OF THE CONSCIOUS SELF	6
Dualism and Religion	6
Varieties of Religious Dualism	8
Naturalism and Identity Theory: An Overview of the Argument of Chapter One	12
Scientific Constraints and the Mind-Body Problem	13
Explaining Consciousness: The Contemporary Mind-Body Problem as a Problem of Vertical Integration	16
The Mind-Body Problem and the Quest for Vertical Integration	17
Positions in the Philosophy of Mind in Light of the Issue of Vertical Integration	18
Identity Theory: A Phenomenally Realist, Pluralist, Neurophilosophical Version	20
Dualism and Consciousness	23
Thomas Nagel: “What Is It Like To Be a Bat?”	23
Saul Kripke and A Posteriori Necessity	24
Joseph Levine and the Explanatory Gap	27
Frank Jackson and the Knowledge Argument	29
David Chalmers and the Hard Problem	31
Dualism in Question	34
The Incomplete Concepts Strategy	35
Subjectivity Naturalized: Neurophilosophical Identity Theory and Phenomenal Realism	42
Naturalism and Phenomenal Realism	43
Subjectivity and Naturalism: Leaving Out ‘What It Is Like’?	47
Consciousness and the Brain: Explaining the Intuition of Distinctness	50
Naturalism and Qualia	53
Paul Churchland’s Neurophilosophical Model of Qualia	55
Naturalistic Anti-Reductionism: Functionalism and the Autonomy of Psychology	59
Functionalism, Multiple Realizability, and the Autonomy of Psychology	60
Multiple Realizability Critiqued	61
Vertical Integration: Beyond Reductive and Non-Reductive Physicalism	66
The Co-Evolution of Psychology and Neuroscience: Reduction and Elimination	66
Beyond Autonomy and Reduction	68
Explanatory Pluralism and Co-Evolution	71
Identity Theory and Explanation	73
Conclusion	77

CHAPTER TWO: TOWARD A NATURALISTIC, BIO-PSYCHOLOGICAL EXPLANATION OF DUALISM	78
Naturalism and the Ubiquity of Dualism	78
Principled Humanist Objections to a Naturalistic, Bio-Psychological Explanation of Dualism	79
Natural Sources of the Dualist Self: Toward a Bio-Psychological Explanation of Dualism	85
Toward a Bio-Psychological Explanation of Dualism: Insights from Neuroscience and Neurophilosophy	86
Defining the Problem of the Conscious Self	88
An Evolutionary Account of the Conscious Self: Antonio Damasio	89
Natural Sources of Religious Dualism I: Selves as Transparent Phenomenal Self Models: Thomas Metzinger	95
Natural Sources of Religious Dualism II: Thomas Metzinger and Olaf Blanke on Out-of-Body Experiences	102
Toward a Bio-Psychological Explanation of Dualism: Insights from Cognitive Science	112
Introducing Cognitive Science	113
Folk Psychology: Agency Detection and Theory of Mind	118
Folk Knowledge and Religious Belief: Introducing the Cognitive Science of Religion	124
Natural Sources of Religious Dualism III: Paul Bloom and Jesse Bering on Folk Psychology and Dualism	126
Natural Sources of Religious Dualism IV: From Dualism to Theism: Over-Attribution of Agency and Design	143
Conclusion	152
CHAPTER THREE: BIO-PSYCHOLOGICAL EXPLANATION AND THE TRUTH OF RELIGIOUS DUALISM	154
Recap: Why Mind-Brain Science Matters for the Study of Religion	154
Dualism and Classical Theism	157
Emergentism and Theism: Philip Clayton	167
Clayton's Account of the Relationship between Theology and Neuroscience	169
Clayton's Account of Emergentism	170
Clayton's Move from Emergentism to Dualist Theism	179
Experiential Religious Dualism: B. Alan Wallace	185
Wallace's Negative Arguments about Naturalism and a Naturalistic Explanation of Consciousness	188
Wallace's Case for Buddhist Religious Dualism	192
CHAPTER FOUR: SHIFTING THE RELIGION AND SCIENCE CONVERSATION: PSYCHOLOGICAL EXPLANATION OF MORALITY AND ITS IMPACT ON METAETHICS AND RELIGION	210
Recap of the Dissertation Argument	210

The Implications of the Naturalistic Conclusions of this Dissertation for the Religion and Science Dialogue	213
Points of Intersection between Religion and Science: Metaphysics and Morals	216
Mind Science, Metaethics and Religion I: Divine Command Ethics and the Origins of Moral Judgment	221
Religion and Morality	221
Religion, Morality, and the Moral-Conventional Distinction	223
The Origin of the Moral-Conventional Distinction	229
The Affective Origins of Morality: Psychopaths and the Moral-Conventional Distinction	229
The Affective Origins of Morality: The Moral Psychology of Disgust	232
Commonsense Moral Objectivism and the Relationship of Religion and Morality	237
Mind Science, Metaethics, and Religion II: Jonathan Haidt and Joshua Greene on the Relationship between Emotion and Reason in Moral Judgment	247
Emotion, Post-Hoc Reasoning, and Morality: Joshua Greene’s Critique of Rationalist Deontology	249
Emotion and Reason in Jonathan Haidt’s Social-Intuitionist Account of Moral Judgment	256
The Biological and Cultural Origins of Moral Intuitions	260
Mind Science, Metaethics, and Religion III: The Implications of Scientific Moral Psychology for Normative Ethics and Religious Practices	263
Scientific Moral Psychology, Virtue Ethics, and Issues of Normative Moral Evaluation	264
A Fruitful Approach to Religion and Science: The Buddhism-Mind Science Dialogue on Meditation, Affective Style, and Emotion	267
Conclusion	275
CONCLUSION	278
BIBLIOGRAPHY	282
VITA	299

ABSTRACT

This dissertation approaches metaphysical and metaethical questions concerning the nature of the human person, the existence and nature of God, and the nature of moral judgment through contemporary neuroscience, cognitive science, scientific moral psychology, and analytic philosophy of mind. Contrary to proposals that seek a harmonious integration of “religion and science” this dissertation argues that contemporary bio-psychological sciences give one ample reason to be skeptical of many of the metaphysical and metaethical claims embedded in religious traditions like Christianity and Buddhism. The first three chapters of the dissertation focus on the metaphysical issue of mind-body dualism, while the fourth chapter addresses closely related metaethical issues concerning the nature and origin of moral judgment.

Chapter One calls into question the truth of dualism and defends a version of psychoneural identity theory as the most adequate solution to the mind-body problem. If dualism is not true, why is it so ubiquitous across cultures? Drawing on research in both neuroscience and cognitive science, Chapter Two offers a naturalistic, bio-psychological explanation of dualism. Chapter Three applies the conclusions of the first two chapters to religious and theological concerns. The link between mind-body dualism and theism is highlighted, and it is argued that the collapse of mind-body dualism renders many versions of theism problematic. Philip Clayton’s attempt to integrate an emergentist doctrine of the human person with a dualist doctrine of God is critiqued, as is B. Alan

Wallace's more experiential defense of religious dualism from a Buddhist perspective. Along with a commitment to dualism, many religious traditions understand moral judgment as something objectively grounded in divine commands and/or reason. Chapter Four argues that research in scientific moral psychology implicating the emotions in moral judgment raises significant questions for theological and reason-based accounts of moral judgment.

INTRODUCTION

This dissertation is concerned with interdisciplinary work at the interface of religion and science, interdisciplinary work at the interface of mind and brain, and their mutual relationship. While questions concerning “religion and science” and questions concerning “mind and brain” are typically pursued independently of one another, the issue of the relationship between these two domains of interdisciplinary inquiry becomes increasingly hard to avoid. Contemporary cognitive science, neuroscience, and philosophy of mind do not merely form topics that can be “related to” or “integrated with” some religious tradition. Rather, contemporary bio-psychological science and philosophy have the potential to radically reconfigure human self-understanding, and along with it, the ways in which religion, science, and their respective claims are understood—or so it will be argued.

Contrary to proposals that seek a harmonious integration of “religion and science,” this dissertation argues that the contemporary bio-psychological sciences give one ample reason to be skeptical of many of the metaphysical and metaethical claims embedded in religious traditions. Many religious traditions presume a commitment to fairly robust versions of metaphysical dualism, while contemporary mind science not only gives one every reason to think that the mind is the brain, it also has begun to offer compelling bio-psychological explanations for the ubiquity of God and soul concepts across cultures. Along with a commitment to metaphysical dualism, many religious

traditions also tend to understand morality as rooted in divine commands or as grounded in an objective transcendent realm that exists independently of human minds. The mind sciences—especially contemporary moral psychology—also give one reason for skepticism about these claims, as the centrality of emotion in moral judgment has been highlighted.

This dissertation pursues both metaphysics and morals as interesting points of intersection between religion and science. The first three chapters focus on the metaphysical intersection, in particular, the mind-body or brain-consciousness problem. The guiding question for the first three chapters is the following: How does contemporary work in the mind sciences impact the mind-body problem, and how does this impact upon religious claims? The fourth and final chapter shifts the religion and science conversation from metaphysics to morals, specifically the question of how work in contemporary scientific moral psychology impacts metaethical questions about the origins of morality and how this impacts both normative ethics and religion.

Chapter One first establishes the link between religion and dualism across a wide range of positions in both the Christian and Buddhist traditions. It then goes on to address issues concerning the impact of science on the mind-body problem. In assessing the impact of the bio-psychological sciences on the mind-body problem, Chapter One presents the mind-body or brain-consciousness problem as a problem of vertical integration between various explanatory levels—phenomenal, psychological, and neurological. The central question it pursues is whether contemporary cognitive science and neuroscience are in a position to explain something like phenomenal consciousness,

or if there are principled reasons to be skeptical about a scientific, especially brain-based, account of consciousness and the conscious self. A number of prominent analytic philosophers of mind—Thomas Nagel, Saul Kripke, Joseph Levine, Frank Jackson, and David Chalmers—have defended property dualist positions on consciousness and thus voiced skepticism about the possibility of brain-based explanations of consciousness. Even more naturalistically oriented philosophers like (early) Hilary Putnam and Jerry Fodor have defended the autonomy of psychology from lower level neuroscientific explanations. Dualist and functionalist arguments present the strongest objections to bio-psychological explanations of consciousness and identity theories of mind. Chapter One criticizes dualist and functionalist arguments and defends an account of bio-psychological explanation and a version of mind-brain identity theory that responds to the worries of both dualists and functionalists.

If dualism is not true, why is it so ubiquitous across cultures? Chapter Two offers a naturalistic, bio-psychological explanation of dualism. Here too, one encounters principled objections to such an explanatory strategy. Just as dualists and functionalists object to bio-psychological explanations of consciousness, humanist scholars object to attempts to offer a scientific account of religious phenomena like dualist beliefs and experiences. It is argued that such principled distinctions between the humanities and the sciences are themselves rooted in a residual dualism, and one that is no longer tenable given the conclusions of Chapter One. Having blocked principled humanist objections, the remaining sections offer a bio-psychological explanation of dualism. Building on the insights of Chapter One, Chapter Two first sketches an account of the conscious self and

draws on the work of neurophilosopher Thomas Metzinger and the neuroscientists Antti Revonsuo and Olaf Blanke to explain two potent sources of dualism: ordinary “out-of-brain” experiences and more extraordinary out-of-body experiences. The chapter then turns to psychological explanations of dualism, drawing particularly on the work of Paul Bloom and Jesse Bering. It is then shown how intuitive dualism combines with a tendency to over-attribute agency and design to generate other immaterial mind concepts like God concepts.

Chapter One calls into question the truth of dualism, while Chapter Two explains the ubiquity of such dualism. Chapter Three applies these conclusions directly to religious concerns. The link between dualism and theism is first addressed. Contemporary classical theists see claims about God and the soul as intimately related. As a result, they have attempted to defend substance dualism as a doctrine about the human person. For classical theists, substance dualism not only affects Christian views on the afterlife, it also has bearing on fundamental issues concerning the existence and nature of God. It is argued that substance dualists are fundamentally right about the link between dualism and theism but wrong to think that substance dualism can be saved as a doctrine about the human person. The remainder of Chapter Three focuses on attempts to integrate dualist religious views with mind science. Philip Clayton’s attempt to integrate an emergentist doctrine of the human person with a dualist doctrine of God is first addressed. Clayton’s version of emergentism is critiqued as a version of property dualism, as is his attempt to move from a downwardly dependent account of human minds to a divine mind that is not downwardly dependent. The chapter concludes with a

consideration of B. Alan Wallace's more experientially focused defense of religious dualism from a Buddhist perspective. Both Wallace's account of consciousness and his attempts to ground religion universal experience are critiqued.

Chapter Four then surveys the conclusions of the dissertation thus far and their bearing on issues concerning the relationship of religion and science. It is argued that claims about dualism are an area of conflict between religion and science, and this conflict is not favorable to traditional religious positions. It is argued that conflict need not be a conversation stopper, nor does it rule out other areas of intersection between religion and science. Chapter Four explores one of these other areas of intersection, moving from issues of metaphysics to morals. The final chapter focuses on the implications of the mind sciences for metaethical questions concerning the nature and origin of moral judgment and the bearing of this on both normative ethics and religion.

CHAPTER ONE
NATURALISM, DUALISM, AND BIO-PSYCHOLOGICAL
EXPLANATION OF THE CONSCIOUS SELF

This chapter first establishes the link between mind-body dualism and religion. It then goes on to show why such dualism has become problematic on scientific grounds and argues that psychoneural identity theory provides the plausible solution to the mind-body or consciousness-brain problem in light of both established scientific conclusions and ongoing scientific research on consciousness.

Dualism and Religion

The concern of this section is simply to point to a link between mind-body dualism and religious traditions like Christianity and Buddhism. Dualist assumptions have a long history in both traditions, and these dualist positions continue to be defended by a number of contemporary theologians and religious thinkers. Such dualist claims are intimately linked to beliefs about the origin, transformation, and destiny of human beings, and, indeed, the entire cosmos. As such, they are hardly tangential claims.

While there are very important differences between religious dualists, they typically share a commitment to two metaphysical theses regarding the relationship of consciousness to the brain. The first thesis, we can call the Correlation Claim (CC): Conscious mental properties are ontologically distinct from physical (e.g., neurological)

properties. Thus the relationship between consciousness and the brain can only be understood in terms of correlation and not identity.

The correlation claim is a claim that all dualists will share. Property dualists typically hold either the position that (1) consciousness is dependent upon the brain (e.g., emergentism), or (2) consciousness and the brain are lawfully correlated (e.g., panpsychism).¹ Substance dualists—those who defend the notion of a soul—hold that this correlation involves no downward dependence; the soul must be directly created by God. Religious dualists can be substance dualists or property dualists about human minds; however, religious dualists—even if they are property dualists and acknowledge the downward dependency of the human mind on the brain—are committed to a more robust metaphysical thesis about consciousness and the brain. This is the claim that there are actual (as opposed to merely possible) cases where consciousness or the conscious self exists without any ties to a nervous system or indeed any physical substrate. We can call this the No Neural Correlates Claim (NNCC): Consciousness or the conscious self is capable of existing without any neural or physical correlates (e.g., the post-mortem soul, God, substrate consciousness).

The remainder of this section cites specific examples of religious dualism. The purpose is to demonstrate that all of these representatives hold both CC and NNCC and also to demarcate important differences between various dualist positions. Religious dualism is not a marginal or idiosyncratic position, nor is it limited to doctrines about the human person and claims about ‘the soul’ in the Christian tradition. Theism itself is often

¹Non-religious defenders of property dualism typically only entertain mind and brain coming apart in *possible* worlds and not the *actual* world.

articulated in dualist terms. As the remainder of this section will illustrate, this commitment to theistic dualism is found not only within classical Christian theism but also in revisionist forms of theism, such as the process panentheism of Philip Clayton. Further, dualism about consciousness is present within versions of Buddhism that eschew a substance ontology of God and the soul.

Varieties of Religious Dualism

Perhaps the most obvious example of religious dualism comes from within classical Christian theism. Within classical Christian theism, God is understood straightforwardly as an immaterial person. Richard Swinburne captures this understanding well when he defines God as “[a] person without a body (i.e. a spirit) present everywhere, the creator and sustainer of the universe....”² Alvin Plantinga nicely spells out some of the implications of this understanding of God: “Given Christian theism, we know that it is at any rate possible that there be immaterial thinking things, since God Himself is such a thing.”³ As Plantinga’s quote illustrates, the notion of God as an immaterial person relates intimately to the notion of human beings possessing an immaterial soul along the lines described in the *Catechism of the Catholic Church*:

The Church teaches that every spiritual soul is created immediately by God... and that it is immortal: it does not perish when it separates from the body at death, and it will be united with the body in the final resurrection.⁴

²Richard Swinburne, *The Coherence of Theism* (Oxford: Clarendon Press, 1977), 2.

³Alvin Plantinga, "Materialism and Christian Belief," in *Persons: Human and Divine*, ed. Peter van Inwagen and Dean Zimmerman (Oxford: Oxford University Press, 2007), 100.

⁴Catechism of the Catholic Church, #366.

Of course, a significant number of Christian philosophers and theologians reject the notion of the soul. As Dean Zimmerman writes:

Christianity is often thought to require a dualistic conception of human persons, according to which each of us has (or perhaps simply *is*) an immaterial soul that survives death and awaits reunion with the body at a general resurrection....For several decades, however philosophers and theologians have been questioning the inevitability of Christian opposition to materialism. Indeed, at present there seem to be more Christian philosophers [and theologians] defending materialism (as a theory about human persons, not about the deity) than dualism—at least in print.⁵

Notice, however, in Zimmerman's quote that "materialism" is embraced as a *local* doctrine about *human* minds, the notion of God as an immaterial person remains strong in such position. This revisionist stance certainly takes much more of a naturalistic perspective than does classical theism; however, its embrace of theism still places it firmly within a dualistic framework. At least one case of an immaterial mind remains within these revisionist forms of theism. The fundamentally dualist nature of these revisionist forms of theism can be seen in the writings of Philip Clayton.⁶ Clayton repudiates both classical theism and substance dualism in favor of process panentheism and emergentism, but the notion of God as immaterial mind floating free from any physical substrate remains strong.

While this position [theistic emergentism] affirms that all mental phenomena in the empirical world are dependent on a biological substrate, it postulates that transcendent mind is not downwardly dependent in this way. This fact accounts for the *ineliminable element of dualism* in the

⁵Dean Zimmerman, "Three Introductory Questions," in *Persons: Human and Divine*, ed. Peter van Inwagen and Dean Zimmerman (Oxford: Oxford University Press, 2008), 2.

⁶Clayton's views about theism are hardly idiosyncratic; indeed, they are representative of the views of the majority of Christian theologians interested in religion and science: Ian Barbour, John Polkinghorne, Arthur Peacocke, Nancey Murphy, John Haught, and Gregory Peterson all defend similar positions.

theistic hypothesis....[T]his move forces the chain of explanation beyond the framework that one otherwise uses to explain mental properties, in so far as it imagines a mind that is *distinct in essence* from the natural order taken as a whole.⁷ (Emphasis added.)

Christian religious dualists thus differ sharply over the issue of dualism as a doctrine about human persons and the implications of that understanding for their doctrine of God.

There are also important interreligious differences between religious dualists over exactly what kind of immaterial minds are thought to exist. Notions of immaterial mind and consciousness also figure prominently in a tradition like Buddhism that lacks both the notion of God and a substantial soul. This can be seen in the recent writings of the fourteenth Dalai Lama and his English translator, B. Alan Wallace.⁸ Wallace, in particular, has highlighted the important role immaterial consciousness plays within Tibetan Buddhism in his recent works on Buddhism and contemporary mind science.

Wallace writes:

Buddhist contemplatives claim that with the achievement of a highly advanced degree of *Samadhi*[sustained attention] known as *samatha*, or meditative quiescence, one gains experiential access to the relative ground state of consciousness known in the Great Perfection (Dzogchen) school of Tibetan Buddhism as the “substrate consciousness” (*alayavijnana*). This, they claim, is the individual stream of consciousness from which the psyche and all the physical senses emerge. According to their findings, the psyche is *conditioned* by the body and its physical interaction with the environment, but it *emerges* from the substrate consciousness.... In the course of a human life, these mental events are conditioned by the brain

⁷Philip Clayton, *Mind and Emergence: From Quantum to Consciousness* (Oxford: Oxford University Press, 2004), 183-184.

⁸Dalai Lama, *The Universe in a Single Atom: The Convergence of Science and Spirituality*; B. Alan Wallace, *The Taboo of Subjectivity: Toward a New Science of Consciousness* (Oxford: Oxford University Press, 2000); *Contemplative Science* (New York: Columbia University Press, 2007).

and environment, but they emerge from and dissolve back into the substrate consciousness. Likewise, these mental events *influence* the brain, body, and the physical environment, but they do not *transform into* those physical phenomena. In short, from this Buddhist perspective, the “hard problem” of how the brain produces subjective mental experience is a false problem, for such experience actually arise from the substrate consciousness.⁹ (Emphasis in original.)

The general nature of religious dualism has been spelled out in terms of CC and NNCC and specific versions religious dualism have been cited. The remaining chapters of this dissertation will show how problematic such claims have become in light of contemporary mind science.

When one reads contemporary scientific and philosophical literature on the conscious self one cannot but be stuck by its distance from the metaphysical and methodological commitments of even the most scientifically informed religious dualists. Bio-psychological explanations of the conscious self are being actively pursued, and most of these research programs are guided by the assumption that mind can be identified with the brain. Mind scientists are well aware of the metaphysical implications of their work and don't hesitate to point them out (much to the irritation of the philosophers and theologians who disagree with them).¹⁰ Part of the burden of this chapter is to show that this tension is real by investigating the tenability of identity theory in philosophy of mind. Scientists and philosophers working at the mind-brain, psychology-neuroscience interface are not captive to the “ideologies” of materialism, reductionism, and scientism.

⁹B. Alan Wallace, *Contemplative Science* (New York: Columbia University Press, 2007), 15-17.

¹⁰Francis Crick, for example, entitles the final chapter of his book on consciousness “Dr. Crick’s Sunday Morning Service.” See Francis Crick, *The Astonishing Hypothesis: The Scientific Search for the Soul* (New York: Simon & Schuster, 1994), 255-263.

Rather work in this area gives us every reason to think that the mind is the brain. This work stands in substantial tension with dualist understandings of the conscious self as it calls into question the mildly dualist Correlation Claim (CC) and the strongly dualist No Neural Correlates Claim (NNCC).

Naturalism and Identity Theory:

An Overview of the Argument of Chapter One

This chapter points to the significant pressures the scientific study of mind puts upon religious dualism. It addresses the mind-body problem and defends a bio-psychological, specifically neurophilosophical, form of identity theory. Its goal is neither definitive proof of mind-brain identity or definitive disproof of dualism. It rather seeks to establish the fundamental plausibility of mind-brain identity theory as the most relatively adequate position in comparison to other understandings of the mind-brain relation.

The argument of the chapter unfolds in the following manner. It begins with the presentation of the mind-body or brain-consciousness problem in light of the issue of bio-psychological explanation and outlines the five canonical positions in the philosophy of mind taken in response to the mind-body problem. In light of this background, this chapter then stakes out its own position on the issue, defending a neurophilosophical form of identity theory. Identity theory, however, has fallen upon some tough times in recent analytic philosophy of mind. This chapter first states major contemporary dualist objections to physicalist understandings of consciousness. Dualism is then critiqued in light of a naturalistic account of subjectivity.

However, there have also been pitched battles between physicalists in the contemporary philosophy of mind, particularly over the issue of “reductive” versus “non-reductive” forms of physicalism. It is argued that much of this debate is rooted in understandings of explanation and intertheoretic reduction no longer tenable in the philosophy of psychology, biology, and neuroscience. A new understanding of identity theory is articulated in light of the notions of mechanistic explanation and explanatory pluralism.

Scientific Constraints and the Mind-Body Problem

With characteristic bluntness, John Searle opens a book on the mind-body problem with the following statement:

The famous mind-body problem, the source of so much controversy over the past two millennia, has a simple solution. This solution has been available to any educated person since serious work began on the brain nearly a century ago, and, in a sense, we all know it to be true. Here it is: Mental phenomena are caused by neurophysiological processes in the brain and are themselves features of the brain. To distinguish this view from the many others in the field, I call it “biological naturalism.” Mental events and processes are as much part of our biological natural history as digestion, mitosis, meiosis, or enzyme secretion.¹¹

This dissertation defends a specific approach to the mind-body problem, a version of identity theory not unlike Searle’s biological naturalism. Searle’s remark is rooted in an understanding of how significantly modern science has reshaped our worldview and our philosophical options. Even prior to the advent of “serious” neuroscience in the twentieth century, scientific advances placed significant constraints on philosophical theorizing.

¹¹John Searle, *The Rediscovery of the Mind* (Cambridge, MA: MIT Press, 1992), 1.

Identity theory becomes a compelling position when one considers three such constraints: call these the physics constraint, the biology constraint, and the neuroscience constraint.

The first constraint, highly significant for the issue of mental causation, we can call the physics constraint. It seems an obvious empirical fact about the world that the conscious self has physical effects in the world (e.g., I go to the refrigerator to get a drink because I am thirsty). Now, physics places important constraints upon our thinking about such activities. Physics operates under conservation laws and the notion that every physical effect has a physical cause (the causal closure of the physical). If the mind is non-physical, one runs into massive difficulties in trying to account for mental causation.

As Owen Flanagan writes:

[T]he principle of conservation of energy requires that the total amount of energy in the universe remain constant, even as it is continually transferred and transformed in and among the myriad systems of causal relations....In order...for physical energy to increase in any system, it has to have been transferred from some other physical system....If we accept the principle of the conservation of energy we seem committed either to denying that the nonphysical mind exists, or to denying that it could cause anything to happen, or to making very implausible ad hoc adjustments in our physics.¹²

The physics constraint has been operative since Descartes, and it is significant that most defenders of dualism defend epiphenomenalism rather than interactionist dualism.¹³

Next, consider the biology constraint. Evolution is a process operating on physical systems; that the human mind is also a physical system makes perfect sense in an evolutionary framework. As Paul Churchland writes:

¹²Owen Flanagan, *The Science of the Mind* (Cambridge, MA.: MIT Press, 1991), 21.

¹³Most significantly Frank Jackson and David Chalmers.

[T]he important point about the standard evolutionary story is that the human species and all of its features are the wholly physical outcome of a purely physical process. Like all but the simplest of organisms, we have a nervous system. And for the same reason: a nervous system permits the discriminative guidance of behavior. But a nervous system is just an active matrix of cells, and a cell is just an active matrix of molecules. We are notable only in that our nervous system is more complex and powerful than those of our fellow creatures. Our inner nature differs from that of simpler creatures in degree, but not in kind.

If this is the correct account of our origins, then there seems neither need, nor room, to fit any nonphysical substances or properties into our theoretical account of ourselves. We are creatures of matter. And we should learn to live with that fact.¹⁴

Finally consider the mind-brain correlations and dependencies described by the contemporary neurosciences. As Patricia Churchland notes:

The degeneration of cognitive function in various dementias such as Alzheimer's disease is closely tied to the degeneration of neurons. The loss of specific functions such as the capacity to feel fear or see visual motion are closely tied to defects in highly specific brain structures in both animals and humans. The shift from being awake to being asleep is characterized by highly specific changes in patterns of neuronal activity in inter-connected regions. The adaptation of eye movements when reversing spectacles are worn is explained by highly predictable modifications in very specific and coordinated regions of the cerebellum and brainstem One of the most metaphysically profound discoveries in this century showed that a human's mental life is disconnected if the two hemispheres of his brain are disconnected.... In careful postoperative studies of the capacities of "the split-brain" subjects, Roger Sperry, Joseph Bogen, and their colleagues found that each hemisphere could have perceptual experiences or make movement decisions independently of the other.... These remarkable results demonstrate that the unity of mental life is dependent on the anatomical connections in the brain itself. This seems reasonable enough on the hypothesis that mental life is activity in the brain.¹⁵

¹⁴Paul M. Churchland, *Matter and Consciousness* (Cambridge, MA: MIT Press, 1984), 21.

¹⁵Patricia Smith Churchland, *Brain-Wise: Studies in Neurophilosophy* (Cambridge, MA: MIT Press, 2002), 43-44.

However, in spite of the explosive growth of the neurosciences in the past century, much work in contemporary analytic philosophy of mind doubts that the brain will have anything significant to tell us about the nature of mind. To understand why this is the case, it is necessary to examine the contemporary mind-body problem in some detail.

**Explaining Consciousness: The Contemporary Mind-Body Problem
as a Problem of Vertical Integration**

The mind-body problem is an ancient problem, receiving one of its early articulations in Plato's *Phaedo*. However, the terms and precise nature of the mind-body problem have changed dramatically as a result of the history of philosophy and the history of science. The *modern* mind-body problem arises from the birth of a new physics in the seventeenth century and receives its classic articulation in Descartes and his early modern contemporaries. Mental causation is the major focus within the modern articulation of the mind-body problem. Mental and physical realms are assumed to be separate, and questions are then raised as to whether and how two such distinct realms could interact with one another.

While mental causation looms large in the current debate, the *contemporary* mind-body—or, more accurately the brain-consciousness—problem focuses on issues of consciousness and its explanation. In the words of Colin McGinn, it focuses on questions of how soggy gray matter gives rise to technicolor phenomenology.¹⁶ The contemporary mind-brain problem is framed not just by considerations concerning modern physics and evolutionary biology, but also by the explosive growth of the sciences of the mind and

¹⁶See Colin McGinn, *The Mysterious Flame: Conscious Minds in a Material World* (New York: Basic Books, 1999).

brain within the twentieth century, and by the question of whether a science of consciousness is in fact possible at all.

The Mind-Body Problem and the Quest for Vertical Integration

The contemporary mind-body problem is perhaps most fruitfully presented as a problem of bio-psychological explanation concerning the relationship between phenomenal consciousness/common sense psychology, scientific psychology, and neuroscience.¹⁷ It is thus intimately related to what has been called the “interface problem” in the philosophy of psychology, the problem of how our common sense or folk psychology (here specifically phenomenal consciousness) interfaces with explanations given by scientific psychology, cognitive science, neuroscience, and other levels in the explanatory hierarchy.¹⁸ As I am presenting it, the contemporary mind-brain problem is a problem concerning the possibility of “vertical integration” or “reflective equilibrium” between phenomenal consciousness, cognitive science, and neuroscience.¹⁹ The neurophilosophical form of identity theory defended in this dissertation is optimistic about the possibility of genuinely explanatory interdisciplinary bridges being built between first-person phenomenology, scientific psychology, and neuroscience, and about the ontological explanation of the conscious self in terms of the brain. However, many

¹⁷For an articulation of the mind-brain problem along these lines, see John Bickle, *Psychoneural Reduction: The New Wave* (Cambridge, MA: MIT Press, 1998).

¹⁸This statement of the interface problem is adapted from Jose Luis Bermudez, *Philosophy of Psychology: A Contemporary Introduction* (New York and London: Routledge, 2005). Bermudez himself tends to place emphasis on distinguishing the interface problem from the mind-body problem, see pp. 13-15 and 35ff.

¹⁹Edward Slingerland discusses the notion of “vertical integration” in Edward Slingerland, *What Science Teaches the Humanities: The Integration of Body and Culture* (Cambridge, UK: Cambridge University Press, 2008); Owen Flanagan discusses consciousness in light of the notion of “reflective equilibrium” in Owen Flanagan, *Consciousness Reconsidered* (Cambridge, MA: MIT Press, 1992).

doubt the possibility of a bio-psychological explanation of consciousness and thus the plausibility of any sort of identity theory of mind. Identity theory has fallen upon tough times, and there has been an eclipse of the brain in much recent analytic philosophy of mind.

Positions in the Philosophy of Mind in Light of the Issue of Vertical Integration

There are typically five canonical positions stated in relationship to the mind-body problem: substance dualism, property dualism, functionalism, identity theory (reductive materialism) and eliminative materialism.²⁰ All of these positions understand the relationship between phenomenal consciousness, scientific psychology, and neuroscience quite differently. An important issue here concerns the criteria by which consciousness is explained. Identity theory is frequently referred to as reductive physicalism, reflecting the fact that identity theory quickly became linked with treatments of intertheoretic reduction in the philosophy of science, specifically Ernest Nagel's model of intertheoretic reduction.²¹ Drawing upon a deductive-nomological (D-N) understanding of scientific explanation, Nagel understood reduction as a relationship between theories (where theories are understood as sets of laws), and argued that reduction should be understood as a *deduction* of the reduced theory from the reducing

²⁰See, for example, Paul Churchland's discussion in *Matter and Consciousness* or John Heil's discussion in John Heil, *Philosophy of Mind: A Contemporary Introduction* (New York and London: Routledge, 2004).

²¹See Ernest Nagel, *The Structure of Science* (New York: Harcourt, Brace and World, 1961).

theory. The notion of reduction as deduction and deductive criteria of explanation have framed much of the discussion in contemporary philosophy of mind.

Substance dualism, property dualism, and functionalism are strong anti-reductive positions and each position in its own way denies that the brain will tell us anything very interesting about the mind. Dualism in principle denies the possibility of vertical integration of phenomenal consciousness, cognitive science, and neuroscience because consciousness just isn't physical. Interactionist substance dualism denies that consciousness and free will can be explained in naturalistic terms; being non-physical they exist outside the causal nexus described by the natural sciences. Interactionist emergentism also places principled limits on the explanatory power of lower level sciences. All forms of interactionism are also premised on a denial of the notion of the causal closure of the physical world or attempt to circumvent the principle. The form of property dualism advocated by most contemporary philosophers maintains that conscious mental properties (qualia) escape the net of lower level functional and physical explanation. Many of the most prominent arguments for dualism today rely on deductive criteria of explanation. It is precisely the failure to be able to deduce consciousness from the physical facts that leads these thinkers to dualist conclusions.

There are also questions raised about vertical integration from within physicalism. As it developed historically, identity theory envisions a smooth Nagel-reduction of common sense psychology and scientific psychology to neuroscience. Functionalism and eliminativism both call into question this understanding of the relationship between

mental states and brain states, and in doing so, also understand the interdisciplinary interface between psychology and neuroscience quite differently.

Functionalism is the most common form of non-reductive physicalism. For functionalists, the essence of a mental state is its causal role linking perceptual inputs, other mental states, and behavioral outputs. According to the functionalist these abstract mental states can be realized in any number of physical systems; mental states are multiply realizable. The implication of this is the autonomy of psychology from lower level neuroscience. For functionalists, the wildly disjunctive nature of mind to brain mappings also rules out deductive entailment.

Identity theory and eliminativism are typically the understandings of mind that take the brain most seriously. Both theories, however, understand vertical integration or reflective equilibrium in terms of reduction. Identity theory is frequently seen as synonymous with reductive materialism or physicalism. The theory anticipates neuroscience smoothly reducing common sense and scientific psychology. Eliminativism is skeptical about a smooth reduction and anticipates a wholesale elimination of common sense psychology and scientific psychology by a matured neuroscience.

Identity Theory: A Phenomenally Realist, Pluralist, Neurophilosophical Version

All of the positions surveyed above are problematic. This chapter defends a *phenomenally realist, pluralist, neurophilosophical* form of identity theory. Originally developed by Paul and Patricia Churchland,²² neurophilosophy is a particular form of

²²For an early statement of neurophilosophy see Patricia Smith Churchland, *Neurophilosophy: Toward a Unified Science of the Mind/Brain* (Cambridge, MA: MIT Press, 1986).

naturalism that seeks to bring the resources of neuroscience to bear on traditional philosophical issues.

[T]he nature of the mind (including the nature of memory and learning, consciousness, and free will) have traditionally been subjects within the purview of philosophy. Philosophers, by tradition, have wrestled with these topics and the work continues. Neurophilosophy arises out of the recognition that at long last, the brain sciences and their adjunct technology are sufficiently advanced that real progress can be made in understanding the mind-brain. More brashly, it predicts that philosophy of mind conducted with no understanding of neurons and the brain is likely to be sterile. Neurophilosophy, as a result, focuses on problems at the intersection of a greening neuroscience and a graying philosophy.²³

The Churchlands are known not only for the advocacy of neurophilosophy, but also for their defense of eliminative materialism. Neurophilosophy and eliminativism have been linked in the minds of many, but the two theses are distinct. This chapter endorses neurophilosophy but not the eliminativist understanding of vertical integration. It will show that one can take the brain seriously without eliminating higher-level mind science and common sense psychology from one's explanatory framework. This chapter also endorses a version of identity theory, but not the form of reductionism with which identity theory is usually associated. The success of neurophilosophy as a program, of course, depends on the neurosciences having a certain degree of explanatory power and for these explanations to be able to yield ontological conclusions. As noted above, dualists and functionalists offer principled objections about the explanatory and ontological implications of the neurosciences.

This chapter thus begins by discussing these dualist and functionalist concerns. It is argued that while neither dualist nor functionalist arguments are compelling, both

²³Churchland, *Brain-Wise: Studies in Neurophilosophy*, 2-3.

theories contain important insights. Dualism is correct to stress the *subjectivity* of consciousness; however, it is argued that naturalism is better positioned to account for subjectivity than dualism. The form of identity theory defended thus embraces *phenomenal realism*.

Phenomenal realism or Inflationism is the view that consciousness is a substantial property that cannot be conceptually reduced or otherwise *philosophically* reduced in non-phenomenal terms....According to most contemporary inflationists, consciousness plays a causal role and its nature may be found empirically as the sciences of consciousness advance. Inflationism is compatible with the *empirical scientific* reduction of consciousness to neurological or computational properties of the brain.... Inflationism accepts the Hard Problem but aims for an empirical solution to it.²⁴

The form of identity theory defended is also pluralist in its understanding of explanation. Functionalism is correct to stress the importance of higher-level cognitive science in the explanation of consciousness, but is wrong in its conclusions about the autonomy of psychology based on multiple realizability arguments. Recent debates between advocates of “reductive” and “non-reductive” physicalism tend to pit psychology and neuroscience against one another in a zero-sum game. The pluralist stance is an attempt to move beyond debates concerning non-reductive and reductive forms of physicalism. This pluralist stance draws upon models of mechanistic explanation and explanatory pluralism in recent philosophy of psychology and philosophy of neuroscience, and argues that the project of vertical integration need not be exclusively understood in reductionist or eliminativist terms.

²⁴Ned Block, "Consciousness, Philosophical Issues About," (2002). Retrieved from: www.nyu.edu/gsas/dept/philo/faculty/block/papers/ecs.pdf (accessed 1/15/09).

Dualism and Consciousness

Dualists argue that consciousness cannot be explained by lower level cognitive science and neuroscience because consciousness just isn't physical. Religious dualists typically regard the entire conscious self as beyond the functional-physical explanatory framework of the sciences. Within contemporary analytic philosophy of mind, the debate focuses more narrowly on the issue of phenomenal consciousness and subjective qualities of experience or qualia. This section provides an overview of the major anti-reductionist arguments based on consciousness in contemporary analytic philosophy of mind.

Thomas Nagel: "What Is It Like to Be a Bat?"

Thomas Nagel was one of the first to draw anti-reductionist, anti-physicalist (but not pro-dualist) conclusions based on consciousness. The terms in which Nagel posed the mind-body problem in "What Is It Like to Be a Bat?" echo through later treatments of the issue. For Nagel, it is consciousness that makes the mind-body problem really intractable.²⁵ According to Nagel, "the fact that an organism has conscious experience at all means, basically that there is something it is like to be that organism."²⁶ Now according to Nagel, if naturalism is to be defended, phenomenal consciousness must also be given a physical account. Nagel argues that this is impossible, however, due to the fact that "every subjective phenomenon is essentially connected to a single point of view, and it seems inevitable that an objective physical theory will abandon that point of view."²⁷

²⁵Thomas Nagel, "What Is It Like to Be a Bat?" in *Philosophy of Mind: A Guide and Anthology*, ed. John Heil (Oxford: Oxford University Press, 2004), 528.

²⁶Ibid., 529.

²⁷Ibid., 530.

All of this has implications for the mind-body problem, for it is a mystery “how the true character of experiences could be revealed in the physical operations of that organism.”²⁸

Phenomenal consciousness simply does not fit typical patterns of reduction in the sciences:

We appear to be faced with a general difficulty about psychophysical reduction. In other areas, the process of reduction is a move in the direction of greater objectivity, toward a more accurate view of the real nature of things. This is accomplished by reducing our dependence on individual or species-specific points of view toward the object of investigation....Experience itself, however, does not seem to fit the pattern. The idea of moving from appearance to reality seems to make no sense here.²⁹

Saul Kripke and A Posteriori Necessity

Work in modal logic by Saul Kripke also challenged the analogy between mind-brain identity and other scientifically discovered identities.³⁰ Early identity theorists argued that psycho-neural identity was similar to others empirically discovered scientific identities like heat = mean molecular kinetic energy or water = H₂O. As these identities are discovered a posteriori, these thinkers also argued that these identities were *contingent*. Significantly, Kripke argued that if identity claims were true at all, they were true in all possible worlds. Regardless if the claim is discovered empirically or not, if it is an identity claim, it holds of necessity. Scientific identities, while discovered a posteriori, like heat = mean molecular kinetic energy or water = H₂O were thus necessary claims.

²⁸Ibid., 533.

²⁹Ibid., 535.

³⁰See Saul Kripke, *Naming and Necessity* (Cambridge, MA: Harvard University Press, 1972); Saul Kripke, "Identity and Necessity," in *Philosophy of Mind: A Guide and Anthology*, ed. John Heil (Oxford: Oxford University Press, 2004).

Kripke realized that such empirically discovered scientific identities certainly don't *seem* necessary, however. Kripke argued that in cases of genuine identity the appearance of contingency could be explained away. Kripke's strategy for doing this again trades on an appearance/reality distinction. The appearance of contingency with scientifically established identities occurs in virtue of the fact that the everyday terms involved in such identities (e.g., water, heat) have their references picked out by contingent properties (i.e., their appearances).³¹ With everyday terms like heat or water reference is fixed by these contingent surface properties rather than the actual nature of the everyday term. Thus water is identified as "the odorless, colorless stuff that falls from the sky" or heat is identified as "the stuff that causes heat sensations." The relationship between water and "the odorless, colorless stuff that falls from the sky" or heat and "the stuff that causes heat sensations" *is* contingent, and it is precisely because these properties fix reference that an illusion of contingency is generated. The contingency lies not with the nature of water (H₂O) but with surface properties that human beings use to pick out water. Once appearance (the odorless, colorless stuff that falls from the sky) is distinguished from reality (H₂O) the appearance of contingency can be explained away.³²

Kripke's insights into the logic of identity had important implications for the mind-body problem, for he argued that there was a profound difference between identity statements like Heat = mean molecular kinetic energy and Pain = c-fibers firing. With purported mind-brain identities, Kripke argued, the strategy of explaining away the

³¹Ibid., 130.

³²Ibid.

contingency used with other scientific identities is not available. We explain away the contingency with other scientific identities by making a distinction between the way heat appears to us and the phenomena itself. Heat appears to us as “the stuff that causes heat sensations” but it is really the motion of molecules. An appearance-reality distinction is not available, in the case of mind-brain identities because the appearance *is* the reality. “The experience has to be *this experience*, and I cannot say that it is contingent property of the pain I now have that it is a pain.”³³ The sensation of pain counts as pain itself and it seems that pain can come apart from c-fibers firing. According to Kripke, the challenge for physicalists is thus quite steep:

The identity theorist who holds that pain is the brain state, also has to hold that it necessarily is a brain state....He has to hold that we are under some illusion in thinking that we can imagine that there could have been pains without brain states. And the only model I can think of for what the illusion might be, or at least the model given by the analogy the materialists suggest, namely, heat and molecular motion, simply does not work in this case. So the materialist is up against a very stiff challenge. He has to show that these things we think we can see to be possible are in fact not possible. He has to show that these things which we can imagine are not in fact things we can imagine. And that requires some very different philosophical argument from the sort which has been given in the case of heat and molecular motion. And it would have to be a deeper and subtler argument than I can fathom and subtler than has ever appeared in any materialist literature that I have read. So the conclusion of this investigation would be that the analytical tools we are using go against the identity thesis and so go against the general thesis that mental states are just physical states.³⁴

³³Ibid., 131.

³⁴Ibid., 132.

The early anti-naturalist sentiments articulated by Nagel and Kripke in the 1970s were shaped into distinctive arguments by Joseph Levine, Frank Jackson, and David Chalmers in the 1980s and 1990s.

Joseph Levine and the Explanatory Gap

While rejecting the metaphysical conclusions of Kripke's modal argument, Joseph Levine amplifies its fundamental epistemological point in "Materialism and Qualia: The Explanatory Gap."³⁵ Kripke's analysis pointed to the difference between statements like "heat is the motion of molecules" and psycho-neural identity statements like "pain is the firing of c-fibers." According to Kripke, if these statements are genuine identities, they must be necessary. In the case of heat, the felt contingency can be explained away; however, the felt contingency with the psycho-neural identity cannot be explained away in the same manner. Levine is concerned with probing why this is the case. He does so in light of the notion of fully explanatory identity. "Fully explanatory identities," according to Levine, are identities that leave nothing crucial out. Thus the identification of heat with the motion of molecules "is explanatory in the sense that our knowledge of chemistry and physics makes intelligible how it is that something like the motion of molecules could play the causal role we associate with heat. Furthermore, antecedent to our discovery of the essential nature of heat, its causal role...exhausts our notion of it."³⁶

³⁵Joseph Levine, "Materialism and Qualia: The Explanatory Gap," *Pacific Philosophical Quarterly* 64 (1983).

³⁶Ibid., 357.

Levine argues that the psycho-neural identity, “pain is the firing of c-fibers,” is also explanatory in this sense, but his point is that in the case of such psycho-neural identities physical and functional analysis does not *exhaustively* explain the concept “pain.”

[T]here is more to our concept of pain than its causal role, there is its qualitative character, how it feels; and what is left unexplained by the discovery of C-fiber firing is *why pain should feel the way it does!* For there seems to be nothing about C-fiber firing which makes it naturally “fit” the phenomenal properties of pain, any more than it would fit some other set of phenomenal properties. Unlike its functional role, the identification of the qualitative side of pain with C-fiber firing...leaves the connection between it and what we identify it with completely mysterious.³⁷

Because qualia cannot be captured in physical or functional terms, it is natural to imagine that these purported identities can in fact come apart:

If there is nothing we can determine about C-fiber firing that explains why having one’s C-fibers fire has the qualitative character that it does—or, to put it another way, if what’s it’s particularly like to have one’s C-fibers fire is not explained, or made intelligible, by understanding the physical or functional properties of C-fibers firings—it immediately becomes imaginable that there be C-fiber firings without the feeling of pain, and *vice versa*. We don’t have the corresponding intuition in the case of heat and the motion of molecules—once we get clear about the right way to characterize what we imagine—because whatever there is to explain about heat is explained by its being the motion of molecules.³⁸

Levine’s argument is an argument from the absence of conceptual analysis.

Because consciousness cannot be exhaustively analyzed in physical or functional terms an explanatory gap will always remain between properties of the brain and conscious mental properties. Importantly, Levine’s argument from the absence of conceptual

³⁷Ibid.

³⁸Ibid., 359.

analysis is an epistemological argument. Physicalism may be true; we just don't understand *how* it can be true. For both Frank Jackson and David Chalmers the lack of deductive entailment between consciousness and the physical implies that conscious mental properties are non-physical.

Frank Jackson and the Knowledge Argument

The lack of entailment from physical facts to phenomenal facts is the focus of a famous thought experiment by Frank Jackson. In "Epiphenomenal Qualia," Jackson explicitly seeks to avoid mere clashes of intuition about qualia by developing an argument whose premises are obvious to all.

There are many qualia freaks, and some of them say that their rejection of physicalism is an unargued intuition. I think they are being unfair to themselves. They have the following argument. Nothing you could tell of a physical sort captures the smell of a rose, for instance. Therefore, physicalism is false. By our lights this is a perfectly good argument.... I must, however, admit that it is weak from a polemical point of view. There are, unfortunately for us, many who do not find the premise intuitively obvious. The task then is to present an argument whose premises are obvious to all, or at least to as many as possible.³⁹

The result is his now famous color scientist Mary:

Mary is a brilliant scientist who is, for whatever reason, forced to investigate the world from a black-and-white room via a black-and-white television monitor. She specializes in the neurophysiology of vision and acquires, let us suppose, all the physical information there is to obtain about what goes on when we see ripe tomatoes, or the sky, and use terms like 'red', 'blue', and so on.... What will happen when Mary is released from her black-and-white room or is given a color-television monitor? Will she *learn* anything or not? It seems just obvious that she will learn something about our world and our visual experience of it. But then it is inescapable that her previous knowledge was incomplete. But she had *all*

³⁹Frank Jackson, "Epiphenomenal Qualia," in *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*, ed. Yujin Nagasawa, Peter Ludlow, and Daniel Stoljar (Cambridge, MA: MIT Press, 2004), 40.

the physical information. *Ergo* there is more to have than that, and physicalism is false.⁴⁰

In a later article, Jackson presents “a convenient and accurate” way to present the argument:

- (1) Mary (before her release) knows everything physical there is to know about other people.
 - (2) Mary (before her release) does not know everything there is to know about other people (because she *learns* something about them on her release).
- Therefore,
- (3) There are truths about other people (and herself) which escape the physicalist story.⁴¹

The argument is an attempt to show that there is no deductive entailment between the physical and the phenomenal; and as a result physical and phenomenal properties must be thought of as distinct. Brian Loar summarizes the argument well:

Consider any phenomenal quality and any physical property however complex. We can know that a person has the physical property without knowing that she experiences the phenomenal quality. And no amount of a priori reasoning or construction can bridge this conceptual gap. That is the intuitive premise. The conclusion is drawn that the phenomenal quality cannot be identical to the physical property. The argument is equivalent to this: since physical and phenomenal conceptions can be connected only a posteriori, physical properties must be distinct from phenomenal properties.⁴²

⁴⁰Ibid.

⁴¹Frank Jackson, "What Mary Didn't Know," in *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*, ed. Yujin Nagasawa, Peter Ludlow, and Daniel Stoljar (Cambridge, MA: MIT Press, 2004), 54.

⁴²Brian Loar, "Phenomenal States (Revised Version)," in *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument* ed. Yujin Nagasawa, Peter Ludlow, and Daniel Stoljar (Cambridge, MA: MIT Press, 2004), 221.

David Chalmers and the Hard Problem

In his 1996 work, *The Conscious Mind*, David Chalmers brings together both modal arguments and knowledge arguments as part of a defense of property dualism.⁴³ Chalmers famously distinguishes between “easy problems” and the “hard problem” in the philosophy of mind. In making this distinction, Chalmers differentiates two concepts of mind: the psychological and the phenomenal. Psychological aspects of mind, including psychological aspects of consciousness may be scientifically perplexing, but they are not particularly philosophically perplexing because the psychological mind can be analyzed in functional and physical terms. Echoing Levine and Jackson, however, Chalmers argues that no such analysis is available for phenomenal consciousness. The “hard problem” is thus the problem of how objective brains give rise to subjective experience. Chalmers argues that consciousness cannot be explained in physicalist terms and then moves to the conclusion that physicalism is false.

In discussing the explanation of consciousness, Chalmers draws upon the notions of supervenience and reductive explanation. Importantly, Chalmers distinguishes between two types of supervenience: logical supervenience and natural supervenience. “B-properties supervene *logically* on A-properties if no two *logically possible* situations are identical with respect to their A-properties but distinct with respect to their B-properties.”⁴⁴ However, Chalmers points out there are other forms of supervenience than

⁴³David J. Chalmers, *The Conscious Mind: In Search of a Fundamental Theory* (Oxford: Oxford University Press, 1996).

⁴⁴*Ibid.*, 35.

logical supervenience. This occurs in cases where there are systematic correlations between two sets of properties in the natural world. This is natural supervenience: “In general, B-properties supervene naturally on A-properties if any two *naturally possible* situations with the same A-properties have the same B-properties.”⁴⁵ Logical supervenience is linked closely to Chalmers’ notion of reductive explanation.

For almost every natural phenomenon above the level of microscopic physics, there seems in principle to exist a *reductive explanation*: that is, an explanation wholly in terms of simpler entities. In these cases, when we give an appropriate account of lower-level processes, an explanation of the higher-level phenomenon falls out.⁴⁶

Logical supervenience is related to reductive explanation in the following manner:

“A natural phenomenon is reductively explainable in terms of some low-level properties precisely when it is logically supervenient on these properties. It is reductively explainable in terms of physical properties—or simply “reductively explainable” when it is logically supervenient on the physical.”⁴⁷ Chalmers argues that consciousness fails to logically supervene upon the physical and thus escapes the net of reductive explanation. Chalmers bases his conclusions about the failure of logical supervenience and reductive explanation on a number of dualist arguments under consideration.

(1) **Absent Qualia:** It is logically possible to conceive of a creature—a philosophical zombie—who is physically and functionally identical to me but lacks conscious experience. Of course, such a being is *empirically* impossible, but the conceivability of zombies shows the failure of logical supervenience and reductive explanation.

⁴⁵Ibid., 36.

⁴⁶Ibid., 42.

⁴⁷Ibid., 47-48.

- (2) **Inverted Qualia:** One can imagine two people who are physically and functionally identical but who have inverted conscious experiences. When I have a red experience my inverted twin has a blue experience and vice versa. As a logical possibility, it seems coherent that experiences could be inverted while physical structure is duplicated exactly.
- (3) **The Knowledge Argument:** Mary cannot know what it is like to see red on the basis of physical facts. Subjective facts escape the physicalist story.
- (4) **Absence of Conceptual Analysis:** No sort of physicalist conceptual analysis—i.e. structural and functional analysis—is possible with consciousness: “For consciousness to be entailed by a set of physical facts, one would need some kind of analysis of the notion of consciousness—the kind of analysis whose satisfaction physical facts could imply—and there is no such analysis to be had.”⁴⁸

On the basis of these arguments, Chalmers argues that scientific explanations of consciousness might well be able to explain psychological properties of the mind and even psychological consciousness, but they fail to explain phenomenal consciousness. In other words, they explain “easy problems” but not the “hard problem.” Writing on a number of different scientific explanations of consciousness, Chalmers argues:

The problems with the models and theories presented here do not lie in the *details*; at least, we have not needed to consider the details in order to see what is wrong with them. The problem lies in the overall explanatory strategy. These models and theories are simply not the *sort* of thing that could explain consciousness.⁴⁹

For Chalmers, the failure of reductive explanation shows that physicalism is false. The failure of logical supervenience and reductive explanation establish consciousness as a further fact, over and above the physical facts of the world. Chalmers thus advocates a form of property dualism:

⁴⁸Ibid., 104.

⁴⁹Ibid., 121.

The arguments do not lead us to a dualism such as that of Descartes, with a separate realm of mental substance that exerts its own influence on physical processes. The best evidence of contemporary science tells us that the physical world is more or less causally closed: for every physical event, there is a physical sufficient cause. If so, there is no room for a mental “ghost in the machine” to do any extra causal work....The dualism implied here is instead a kind of *property* dualism: conscious experience involves properties of an individual that are not entailed by the physical properties of that individual, although they may depend lawfully on those properties. Consciousness is a *feature* of the world over and above the physical features of the world...[T]here are properties of individuals in the world—the phenomenal properties—that are ontologically independent of physical properties.⁵⁰

Dualism in Question

The ambitious nature of the arguments of Kripke, Jackson, and Chalmers should not be missed. They seek to establish the truth of dualism and the falsity of physicalism by a priori means alone. The sections that follow mount a critique and response to dualism. Dualist arguments are first critiqued in light of the incompleteness of our concepts concerning consciousness and the brain. Dualist arguments based on a priori conceptual analysis assume that our concepts of both consciousness and brain processes are complete and that the identity conditions for both sensations and brain processes are fully understood. If our knowledge of such concepts and their mutual relationships is not complete, this casts significant doubt on dualist arguments. Conceptual incompleteness cannot establish physicalism, but it does give substantial reasons to be skeptical of dualist arguments.

Even though their conclusions are false, much can be learned from dualist arguments. Dualist arguments do establish the importance of taking conscious

⁵⁰Ibid., 125.

subjectivity seriously and the inadequacy of any approach that does not do so. However, contrary to dualist claims, one can fully affirm subjectivity while remaining a card-carrying physicalist. Thus a naturalistic account of subjectivity is developed following the critique of dualism. It is argued that a naturalistic account enables one to take subjectivity seriously while not ruling out explanations at the levels of scientific psychology and neuroscience.

The Incomplete Concepts Strategy

The methodological and ontological problems for neurophilosophy and identity theory seem enormous indeed. Neurophilosophy and identity theory are predicated on the possibility of some sort of “reflective equilibrium” being established by phenomenal consciousness, cognitive science, and neuroscience.⁵¹ Modal and knowledge arguments seem to call into question the entire ontology of physicalism. If consciousness is, as Chalmers claims, a deep further fact about the world, the most one will be able to expect from cognitive and neuroscientific theories of consciousness are lawful correlations. Consciousness just isn’t the sort of thing that can be explained scientifically. However, there are ample reasons to doubt the property dualism proposed by Kripke, Jackson, and Chalmers.

The arguments of Kripke, Jackson, and Chalmers are rooted in a certain conception of philosophy, a conception that appeals to hypothetical a priori reasoning in order to establish metaphysical truths about the fundamental nature of reality. Note that dualism is already at odds with the physics, biology, and neuroscience constraints

⁵¹See Flanagan, *Consciousness Reconsidered*, 11ff.

discussed earlier. It is surely wrongheaded simply to dismiss such arguments, but a certain amount of skepticism is more than justified. Dualism does not cohere well with our current knowledge of physics, evolutionary biology, and neuroscience. There are also ample reasons to be skeptical of dualist arguments in light of inductive evidence from the history of science. Indeed, the history of modern science reads as a long litany of once indisputable explanatory gaps being closed. Seventeenth century physics, particularly Newton's physics, closed the explanatory gap between sublunary and superlunary realms so central in Aristotelian physics. While Immanuel Kant confidently declared that there would never be a Newton of a blade of grass, nineteenth and twentieth century biology closed the explanatory gap between life and mechanistic physical processes. The point of this section is not to establish physicalism or naturalism, but rather to cast significant doubt about the rather robust claims advanced in arguments by Kripke, Jackson, and Chalmers.

The fundamental issue at stake is whether our concepts of phenomenal consciousness and our concepts of neurophysiological states are well developed enough to draw dualist conclusions about them. Philosophers like Robert Van Gulick, Patricia Churchland, and Thomas Polger have voiced skepticism about the metaphysical conclusions drawn by Kripke, Jackson, and Chalmers. Note that the debate revolves around empirical concepts and a posteriori necessity, and as Kripke and Chalmers both acknowledge, it is possible to be misled by conceivability arguments concerning a posteriori necessity. We might think that heat can come apart from mean molecular kinetic energy or that water can come apart from H_2O , but it really is not possible. Such

thoughts might be *conceivable* in some sense, but they are not really *possible*.

Significantly, for Kripke and Chalmers, we can explain away the appearance of contingency with other identities, but this strategy is not available with consciousness and brain processes. However, examples like heat and water come from largely completed sciences like chemistry and physics.

The disanalogy between purported mind-brain identities and other empirically discovered identities might be due to the fact that neuroscience is simply a very young science. Debates concerning vitalism in biology provide an instructive example. Life was thought to be a fact “over and above” the physical structures and functions of an organism; however, thanks to nineteenth and twentieth century advances in the life sciences, we know that life is not a deep further fact—it just is certain underlying mechanical processes. Dualists have a ready reply to the vitalism analogy. It is possible to conceptually analyze the concept “life” in functionalist terms while it is impossible to do this with phenomenal consciousness. Note, however, that this example of conceptual analysis is applied post facto to a largely completed science. It was by no means evident in the nineteenth century that such a conceptual analysis was available.

Robert Van Gulick has highlighted these issues in what he calls an argument from “parallel conceptual inadequacy.” He argues that one could imagine a nineteenth century vitalist biologist deploying the following argument:

1. I can conceive of creatures that are just like actual creatures (say actual cats) in all physical respects but that have no ability to reproduce.

2. Therefore the ability to reproduce does not logically supervene on a creature's physical structure.⁵²

With the benefit of nineteenth and twentieth century science, however, we know that the vitalist conclusion is just dead wrong. The vitalist lacks both an adequate concept of reproduction and an adequate concept of the total physical structure of the organism, as well as an adequate theory of how the two might fit together.

The vitalist might have conjoined *his concept* of the total physical structure of a cat with the negation of *his concept* of the ability to reproduce without generating any a priori contradictions, but given the radical incompleteness of his concepts vis-à-vis the natures of the two phenomena to which he applied them, nothing really follows regarding what relations of logical (or metaphysical) possibility might hold between these phenomena themselves.⁵³

The implications for consciousness become clear. Our concepts of consciousness and neurophysiology are simply too primitive to draw any putative conclusions regarding their non-identity. Now, as Van Gulick acknowledges, the vitalist analogy also leaves the naturalist with a hefty IOU. The nineteenth century materialist seeking to reply to the vitalist lacked adequate conceptual tools by which to do so, but “carrying the burden of a large promissory note is not the same as having been refuted.”⁵⁴

Patricia Churchland has raised similar reservations about the adequacy of our concepts of consciousness. Churchland draws attention to the fact that conceptual boundaries are constantly being re-drawn in the history of science. Fire provides an

⁵²Robert Van Gulick, "Conceiving Beyond Our Means: The Limits of Thought Experiments," *Toward a Science of Consciousness 3* (2003). Retrieved from: http://cognet.mit.edu/posters/TUCSON3/Van_Gulick.html (accessed 11/17/08).

⁵³Ibid.

⁵⁴Ibid.

instructive example of conceptual re-classification.⁵⁵ Pre-theoretically “fire” included any phenomena that emitted heat and light. Thus burning wood and the sun were thought to share the same “fiery” essence; burning wood, rusting, and bodily metabolism were seen as utterly distinct phenomena. From the benefit of modern physics and chemistry, we now know that burning wood and the sun are governed by two entirely different processes: fast oxidation and nuclear fusion; we also know that burning wood, rusting, and bodily metabolism are in fact examples of the same process: oxidation. The moral of all this for Churchland is that explicit definitions only become available fairly late in the game, as the science that embeds them firms up and matures. In light of this, Churchland voices skepticism regarding Chalmers’s hard-and-fast division of consciousness into hard and easy problems.

My suspicion with respect to the Hard Problem strategy is that it seems to take the class of conscious experiences to be much better defined than it is. The point is, if you are careful to restrict your focus to the prototypical cases, you can easily be hornswoggled into assuming the class is well-defined. As soon as you broaden your horizons, troublesome questions about fuzzy boundaries, about the connections between attention, short term memory and awareness, are present in full, what-do-we-do-with-that glory.⁵⁶

Conceptual inadequacy thus provides a clear response to both the modal argument and the knowledge argument. With regard to Kripke’s modal arguments, conceptual inadequacy becomes a way of explaining away contingencies without appealing to appearance/reality distinctions. In order to know whether a purported identity holds or

⁵⁵Churchland, *Brain-Wise: Studies in Neurophilosophy*, 129ff.

⁵⁶Patricia Smith Churchland, "The Hornswoggle Problem" (1996). Retrieved from: <http://www.wm-johnston.co.uk/philosophy/hornswoggle.htm> (accessed 11/21/08).

doesn't hold, one must be aware of its identity conditions. However, at present, we do not know the identity conditions of either sensations or brain processes. This might in fact explain the appearance of contingency. That is all a response to Kripke requires. As Thomas Polger notes:

I argue that, in our current ignorance of identity conditions of both sensations and brain processes, as far as we know sensations and brain processes may be identical. This is enough...to explain why we might think that the relationship between sensations and brain processes is contingent even if it is not...Remember Kripke's challenge to the naturalist: if we cannot explain the appearance of contingency between the objects of putative identity, then we must deny the identity; but if we can explain the appearance of contingency, then it is possible that the items are identical.⁵⁷

The conceptual inadequacy strategy also provides an initial response to the knowledge argument. At the very least, it suggests that caution is appropriate when drawing strong metaphysical conclusions based on hypothetical reasoning about incomplete concepts. As

Robert Van Gulick observes:

The hypothetical epistemic situation is so unlike the real world contexts in which we operate that it is difficult to have uncontroversial intuitions about it. Can the knowledge argument's defenders make the case that their imagined scenario of Mary's postrelease surprise is the only a priori acceptable one? Can they assume that the alternative imagined by its critics—can be ruled out as impossible by pure a priori reflection on our concepts? It is a big assumption to make, and one can already feel the counterbalancing of plausibility between the argument's assumptions and its radical conclusion. The argument's need for a decisive intuitive judgment in favor of its preferred scenario may seem a thin reed on which to hang so momentous a metaphysical result.⁵⁸

⁵⁷Thomas Polger, *Natural Minds* (Cambridge, MA: MIT Press, 2004), 68.

⁵⁸Robert Van Gulick, "So Many Ways of Saying No to Mary," in *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*, ed. Yujin Nagasawa, Peter Ludlow, and Daniel Stoljar (Cambridge, MA: MIT Press, 2003), 373.

Daniel Dennett, in particular, has exploited this style of response to the knowledge argument, arguing that it is just as plausible to assume that Mary doesn't learn anything upon her release, given her *exhaustive* knowledge of all the physical facts. Dennett has us imagine Mary being presented with a blue banana upon her release:

And so, one day, Mary's captors decided it was time for her to see colors. As a trick, they prepared a bright blue banana to present as her first color experience ever. Mary took one look at it and said 'Hey! You tried to trick me! Bananas are yellow, but this one is blue!' Her captors were dumbfounded. How did she do it? "Simple," she replied. "You have to remember that I know *everything*—absolutely everything—that could ever be known about the physical causes and effects of color vision. So of course before you brought the banana in, I had already written down, in exquisite detail, exactly what physical impression a yellow object or a blue object would have upon my nervous system. So I already knew exactly what *thoughts* I would have (because, after all, the 'mere disposition' to think about this or that is not one of your famous qualia is it?). I was not in the slightest surprised by my experience of blue....I realize it is *hard for you to imagine* that I could know so much about my reactive dispositions that the way blue affected me came as no surprise. Of course it's hard for you to imagine. It's hard for anyone to imagine the consequences of someone knowing everything physical about anything!"⁵⁹

Dennett's point is not that his twist on the knowledge argument establishes physicalism, it is rather that one should not mistake "a failure of imagination for an insight into necessity."⁶⁰

Considerations of conceptual inadequacy do not establish physicalism, but they certainly raise suspicion about the dualist conclusions drawn by Kripke, Jackson, and Chalmers. The conceptual inadequacy strategy is helpful in refuting arguments maintaining *in principle* that one cannot identify consciousness and brain processes;

⁵⁹Daniel C. Dennett, *Consciousness Explained* (Boston: Little, Brown, 1991), 399-400.

⁶⁰Ibid.

however, it isn't a particularly constructive strategy. Indeed, considerations of conceptual inadequacy lead Colin McGinn to maintain that there can *in principle* be no constructive account of the mind-body problem.⁶¹ As a naturalist, McGinn is convinced that such a concept must exist; however, knowledge of such a concept simply exceeds the cognitive abilities of human beings. McGinn's anti-constructive naturalism or mysterian position is not warranted however.⁶²

If one wants to be a *constructive* naturalist, one owes more of an account of the place of subjectivity in the natural world than the conceptual inadequacy strategy provides. The knowledge argument is a failure as an argument for dualism, but it is of great heuristic value for physicalists. It clues one into what an adequate account of physicalism must amount to. In particular, it points to the importance of *taking subjectivity seriously*. Physicalism is thought to preclude a robust account of subjectivity, but this is not the case. The next section examines issues of physicalism and subjectivity in response to the knowledge argument.

Subjectivity Naturalized: Neurophilosophical Identity Theory and Phenomenal Realism

The incomplete concepts strategy effectively blocks arguments *for* property dualism; however, this is just a first pass. The refutation of property dualism certainly doesn't suffice to establish naturalism as a compelling position. More needs to be said about naturalism and its relationship to subjectivity. The next sections will argue that

⁶¹See McGinn, *The Mysterious Flame: Conscious Minds in a Material World*.

⁶²“Anti-constructive naturalism” and “mysterianism” are both terms Owen Flanagan uses to describe McGinn's position. See Flanagan, *Consciousness Reconsidered*.

naturalism is able to respond to the concern that naturalism “leaves out what it is like;” it argues in fact, that naturalism provides the most compelling account of subjectivity currently available.

Naturalism and Phenomenal Realism

Frank Jackson’s knowledge argument proves particularly helpful as a point of departure in exploring issues concerning phenomenal consciousness and naturalism. Let us first review Jackson’s “convenient and accurate” way of displaying the knowledge argument:

- (1) Mary (before her release) knows everything physical there is to know about other people.
- (2) Mary (before her release) does not know everything there is to know about other people (because she *learns* something about them on her release).
Therefore,
- (3) There are truths about other people (and herself) which escape the physicalist story.⁶³

The two responses to the argument seem to be either agreeing with Jackson that Mary learns something on her release, which leads to property dualism, or rejecting premise (2) and denying that Mary learns anything. But, might one be able to admit that Mary does indeed learn something in a fairly robust sense and still remain a physicalist? In approaching this question, Ned Block’s distinction between inflationist and deflationist forms of physicalism is helpful.⁶⁴ Deflationists hold that consciousness can be conceptually analyzed in non-phenomenal terms, thus Mary does not learn anything upon

⁶³Jackson, "What Mary Didn't Know," 54.

⁶⁴See Block, “Consciousness, Philosophical Issues About.”

her release. Inflationist physicalism, on the other hand, is a form of phenomenal realism. Inflationists argue that consciousness cannot be *conceptually* reduced or otherwise philosophically reduced in non-phenomenological terms, but that consciousness is nonetheless physical.⁶⁵

Jackson's argument equates physical knowledge with third person physical concepts that refer to physical properties of the brain.⁶⁶ These physical science facts are taken to be exhaustive of the physical. Mary has knowledge of all the physical concepts and these concepts refer to all the physical properties of the brain. Because Mary gains knowledge, Jackson argues that consciousness must involve non-physical properties. Subjective facts are thus non-physical facts. However, physical knowledge might involve first person phenomenal concepts as well as third person physical concepts. Both first person phenomenal concepts and third person physical concepts refer to physical properties in the brain, but phenomenal concepts may only be acquired through experience. Mary might have knowledge of all the physical properties of the brain, but she knows these facts only in virtue of physical concepts. Because she has never seen red, she lacks the phenomenal concepts that also refer to the same physical properties in the brain. In this sense, "physical knowledge" might extend beyond objective physical science fact and include states that obtain *in virtue of* the physical science facts. In this sense, physical knowledge might be both objective and subjective, and one has an account of how Mary might gain knowledge without this involving any non-physical

⁶⁵Ibid.

⁶⁶The following discussion is indebted to arguments developed by David Papineau, Brian Loar, Robert Van Gulick, and Owen Flanagan.

facts or non-physical properties. Mary's acquisition of new knowledge may be purely epistemological rather than metaphysical.

In this account, seeing red is an entirely physical experience. Prior to her release, Mary knows all about this experience from a third person, objective perspective, but her red channels have never been activated. Upon her release the perceptual/phenomenal concept is triggered in Mary's mind by the basic biological mechanisms that subserve normal color perception. Now an experience of red is not the language of physics. But an experience of red is a physical event in a suitably hooked-up system, a subjective physical fact. As Owen Flanagan notes:

Completed physics, chemistry, and neuroscience, along with a functional-role description, will explain what an experience of red is, in the sense that they will explain how red experiences are realized, what their functional role is, and so on. But no linguistic description will completely *capture* what a first-person experience of red is like. That is only captured in the first-person. You have to be there...[W]e need to beware of the temptation to think that for physicalism to be true, the basic physical sciences must be able to *capture* all truths. This is stronger than requiring that physicalism be true; that is, it is stronger than requiring that everything that happens is physical. Physicalism can be true in this sense without being able to explain everything, let alone capture everything in the languages of the basic physical sciences.⁶⁷

The issue at stake is really epistemological and semantic and not ontological. Third person physical concepts do not suffice for picking out every state of affairs in the world. Physicalism/naturalism is not scientism. Everything that is real is physical in the sense that it is in some way or another realized by underlying physical processes or structures.

⁶⁷Owen Flanagan, *Consciousness Reconsidered*, 100.

But the universality of physical realization is compatible with our epistemic and theoretical need for a diversity of means of representing and modeling reality.⁶⁸

One can remain a card carrying physicalist and yet admit that the knowledge argument reveals something important about consciousness: namely, that we have two ways of thinking about it with third-person person physical science concepts and first-person phenomenal concepts.⁶⁹ The key point is that the dualism is not at the level of substances or properties but rather the concepts of properties. H₂O and water both name the same properties but they are distinct concepts. In Fregean terms, the reference is the same but the sense is different. Prior to her release, Mary knew all about the subjective experience of red via physical concepts, after she left the room she acquired a phenomenal concept of the same property. Mary thus indeed learns something new. She learns new ways to *think* about consciousness that she did not have before.

Responses are thus available for all of the arguments that naturalism cannot *in principle* explain consciousness. Because two concepts are available for thinking about consciousness, we can be led into thinking that the physical and the phenomenal can be separated—as in the case of philosophical zombies—when, in fact they cannot. Philosophical zombies may be conceivable but they are not *possible*. Mary does indeed learn something upon her release, but as we have seen, that is no brief against physicalism. The world is exhausted by physical facts but certain physical facts—subjective physical facts—can only be had through experience. The argument is really an

⁶⁸Van Gulick, "So Many Ways of Saying No to Mary," 381.

⁶⁹See Papineau, *Thinking about Consciousness*, for perhaps the fullest exposition of this position.

argument for empiricism and the limits of discursive knowledge rather than any sort of argument against physicalism. None of the arguments for property dualism are successful. But naturalism is more than anti-dualism. More needs to be said about the place of the conscious self in the natural world.

Subjectivity and Naturalism: Leaving Out ‘What It Is Like’?

If the arguments in the previous sections have been successful, it is clear that dualist arguments in the philosophy of mind are not successful. But what about the claim that naturalistic accounts of consciousness “leave out what it is like”? As Thomas Nagel objected, “[A]ny shift to greater objectivity—does not take us nearer to the real nature of the phenomenon it takes us farther away.”⁷⁰

Owen Flanagan has offered an important critique of Nagel’s line of thought, particularly with regard to ambiguities in the use of the term “real nature.” “Real nature” might mean (1) the way things seem for a particular person (the sense probably intended by Nagel), (2) what is going on with the cognitive system as a whole.⁷¹ In terms of the first sense, Flanagan acknowledges that it is true that a purely objective account cannot capture points of view or the way things seem to someone. However, it should be noted that such a critique equates physicalism with *objective physical facts*, but, as the previous sections of this chapter have illustrated objective physical facts do not exhaust physicalism, for there are subjective physical facts as well.

⁷⁰Ibid., 174.

⁷¹Owen Flanagan, *Consciousness Reconsidered* (Cambridge, MA: MIT Press, 1992), 90.

Nagel makes precisely the same mistake as Jackson. Objective analysis cannot capture points of view, but that is no brief against naturalism, for the doctrine of naturalism includes both objective physical facts and subjective physical facts. As Flanagan notes naturalism is perfectly well positioned to explain “why only you can capture what it’s like to be you: “Only your sensory receptors and brain are properly hooked up to each other and to the rest of you so that what is received at those receptors accrues to you as your experiences.”⁷² This same point has been made by David

Papineau:

Despite the plausibility of Nagel’s line of argument, I think that physicalism can meet the challenge he poses....For a start, we should immediately concede that there is one sense in which we human beings are indeed cut off from the facts of bat experiences. We do not *have* echolocatory experiences, whereas bats do. In this sense it is undoubtedly true that we ‘lack access to’, ‘cannot appreciate’, or whatever phrase you prefer, the ‘subjective reality’ of bat experience. But this observation in itself clearly yields no argument against physicalism. For physicalists are just as well placed as anybody else to explain this difference between bats and humans. Physicalists think that conscious experiences are identical with certain physical events in the brain. So physicalists can say that the difference between bats, who have echolocatory experiences, and humans, who do not, is simply that certain physical events, namely those which constitute echolocatory experiences, occur in bats, but not in humans. In this sense the physicalist can happily agree that bats have access to experiences which humans cannot appreciate.

The point is central to the physicalist view of conscious experience. Physicalism does not deny that there are conscious experiences, nor if you wish, ‘that it is like something to have them.’ The claim is only that this is nothing different from what it is to *be* a physical system of the relevant kind.⁷³

⁷²Ibid., 94.

⁷³David Papineau, *Philosophical Naturalism* (Oxford: Blackwell, 1993), 105-106.

In the second sense of “real nature”—referring to the cognitive system as a whole—a naturalist account does not take one away from the “real nature” of the phenomena at all, anymore than an analysis of water as H₂O takes one away from the “real nature” of water. Precisely because subjective facts are physical facts means they can be analyzed in third person scientific terms. As Flanagan notes: “The important point is this: there is absolutely no reason why a naturalist cannot both acknowledge the existence of subjectivity and view getting an accurate description of it as part of the overall project of understanding human nature....It is crucial to see that description at the phenomenal level is something that can be provided from the first-person point of view and the third-person point of view.”⁷⁴ Subjective physical facts and objective physical facts are merely two sides of the same coin. But what about the complaint that objective physical facts do not “capture” the “true” character of phenomenal experience? This complaint misses the mark. Third person descriptions explain consciousness, they do not create consciousness any more than a theory of digestion creates digestion.

As Flanagan writes:

[W]e do understand how physicalism can be true. It can be true if every mental event is realized in the brain. Those of us who believe that all mental events, conscious and unconscious, are tokened in the brain do not believe that the theory that eventually explains *how* they are tokened will capture “the true character of the experiences” as experiences. The whole idea that the qualitative feel of some experience should reveal itself in a theoretical description of how that experience is realized fails to acknowledge the abstract relation between any theory and the phenomena it accounts for.... [T]he naturalist is the first to accept that a particular

⁷⁴Flanagan, *Consciousness Reconsidered*, 92.

realization will be an experience only for the agent who is causally connected to the realization in the right sort of way.⁷⁵

Consciousness and the Brain: Explaining the Intuition of Distinctness

Mind-brain identity is thus similar in ways to identities like $H_2O = \text{water}$, but an intuition of distinctness remains with mind-brain identities that does not with other identities. There is, in fact, growing evidence from developmental and cognitive science that mind-body dualism is the cognitive default of human beings;⁷⁶ this issue will be explored in much greater detail in Chapter Three. However, other reasons for the “intuition of distinctness” can be found in the anatomy and physiology of our brains, in the structure of our conscious experience, and in our concepts of conscious experience. Mind and brain seem distinct because all of our experiences are “out-of-brain” experiences. There are several factors behind the “out-of-brain” nature of conscious experience. Our subjective experiences don’t seem to be brain experiences. In terms of basic anatomy and physiology, the cranium prevents us from observing the workings of our own brains in the way that we can observe the workings of our muscles. As Rodolfo Llinas has observed, this is no small thing:

If we had an external skeleton, the concept of how movement is generated might be just as incomprehensible to us as is the concept of thinking or mindness. Having an internal skeleton, means that we become quite aware of our muscles from birth. We can see their movement and feel their contractions and clearly understand, in a very intimate way, their relation

⁷⁵Ibid.

⁷⁶See Jesse M. Bering, "The Cognitive Psychology of Belief in the Supernatural," *American Scientist* 94, no. 2 (2006); Jesse M. Bering, "The Folk Psychology of Souls," *Behavioral and Brain Sciences* 29, no. 5 (2006); Jesse M. Bering, and David F. Bjorklund, "The Natural Emergence of Reasoning About the Afterlife as a Developmental Regularity," *Developmental Psychology* 40, no. 2 (2004); Paul Bloom, *Descartes' Baby: How the Science of Child Development Explains What Makes Us Human* (New York: Basic Books, 2004).

to the movement of our different body parts. Unfortunately, we do not have such direct knowledge concerning the workings of our brain. Why not? Because from a cerebral mass point of view, we are crustaceans—our brains and spinal cord are covered by exoskeleton!

If we could observe or feel the brain at work, it would be immediately obvious that neuronal function is as related to how we see, interpret, and react, as muscle contractions are related to the movements we make.⁷⁷

Access to the brain is limited in an even more significant manner due to the fact that the brain has no sensory apparatus directed toward itself; “the brain is not about itself.”⁷⁸ As Todd Feinberg observes:

As Globus puts it, the brain does not “represent in any way its own structure to the subject.” There is no way that the subject can become aware of his own neurons “from the inside.” They can be known only objectively from the “outside.” We have already seen that there is no “inner eye,” no inner homunculus watching the brain itself, perceiving its own neurons, no “brain-skin” which feels the neurosurgeon’s knife. When I test a patients’ pinprick sensitivity by applying a pin to the hand, and ask them to localize where on the body the sensation is, no one has ever pointed to their head. Conscious neural activity refers to things, not to the brain itself. Conscious neural states are *about* things, not about the neurons themselves.⁷⁹

Precisely because conscious neural states are about things and not about the neurons themselves it is very easy to confuse the qualities of conscious experience with the qualities of the objects experienced. U.T. Place dubbed this the “phenomenological fallacy.” Once the distinction is made, and once we realize that consciousness has a depth structure hidden from introspective access, identity theory becomes much more plausible.

⁷⁷Rodolfo Llinas, *I of the Vortex: From Neurons to Self* (Cambridge, MA: MIT Press, 2001), 4.

⁷⁸Todd Feinberg, *Altered Egos: How the Brain Creates the Self* (Oxford: Oxford University Press, 2001), 147.

⁷⁹Ibid.

Finally, the nature of our concepts of consciousness easily fuels the intuition of distinctness. There are two ways of thinking about consciousness: with third-person physical concepts and with first-person phenomenal concepts. As thinkers like David Papineau have pointed out, phenomenal concepts have the peculiar feature of *resembling* the conscious properties they are referring to. When we deploy a phenomenal concept either through imagination or introspection we recreate the experience itself; exercising a phenomenal concept feels like having the experience itself. No such thing happens with material concepts of the same experience. These two ways about thinking about concepts—via phenomenal concepts and via material concepts—can very easily lead us to believe that they are referring to entirely different things.

There is a sense in which material concepts do ‘leave out’ the feelings. They do not *use* the experiences in question—they do not activate them, by contrast with phenomenal concepts which do activate the experiences. But it simply does not follow that material concepts ‘leave out’ the feelings in the sense of failing to *mention* them. They can still refer to the feelings, even though they don’t activate them.

After all, most concepts don’t use or involve the things they refer to. When I think of being rich, say, or having measles, this doesn’t in any sense make me rich or give me measles. In *using* the states they mention, phenomenal concepts are very much the exception. So we shouldn’t conclude on this account that material concepts, which work in the normal way of most concepts, in not using the states they mention, fail to refer to these states.

This then offers a natural account of the intuitive feeling that conscious experiences must be distinct from any material states. This feeling arises because we have a special way of thinking about our conscious experiences—namely by using phenomenal concepts. We can think about our conscious experiences using concepts to which they bear a phenomenal resemblance. And this then creates the fallacious impression that other, material ways of thinking about those experiences fail to refer to the felt experiences themselves.⁸⁰

⁸⁰Papineau, *Thinking About Consciousness*, 171.

But this is simply a use-mention fallacy: because third-person ways of thinking about conscious experience do not *use* versions of those conscious experiences, we conclude falsely that these third person ways of thinking about consciousness do not mention those conscious experiences, but only physical states.⁸¹ Papineau calls this fallacy the “antipathetic fallacy.”

Ruskin coined the phrase ‘pathetic fallacy’ for the poetic figure of speech which attributes human feelings to nature....I am currently discussing a converse fallacy, where we refuse to recognize that conscious feelings inhere in certain parts of nature, namely the brains of conscious beings.⁸²

Naturalism and Qualia

Subjectivity thus has no dire consequences for physicalism. There are no principled reasons why qualia cannot be explained in a naturalistic perspective. However, an opposite problem presents itself within certain forms of naturalism. Some philosophers of a naturalistic bent have argued that the entire notion of “qualia” should be jettisoned.

Daniel Dennett proposes just such a move in *Consciousness Explained*:

When your kite string gets snarled up, in principle it can be unsnarled, especially if you’re patient and analytic. But there’s a point beyond which principle lapses and practicality triumphs. Some snarls should just be abandoned. Go get a new kite string. It’s actually cheaper in the end than the labor it would take to salvage the old one, and you get your kite airborne again sooner. That’s how it is, in my opinion, with the philosophical topic of qualia, a tormented snarl of increasingly convoluted and bizarre thought experiments, jargon, in-jokes, allusions to putative refutations, “received” results that should be returned to sender, and a bounty of other sidetrackers and time-wasters.⁸³

⁸¹Papineau, *Philosophical Naturalism*, 116-117.

⁸²Ibid., 116.

⁸³Dennett, 369.

This section offers a more nuanced assessment of qualia. We can talk intelligently about qualia while avoiding either dualist or eliminativist conclusions about them. A naturalistic account that aims to take subjectivity seriously must address the issue of qualia or the way things seem. Owen Flanagan has offered a number of important insights about qualia from a naturalistic perspective. Flanagan importantly distinguishes between two ways in which qualia can be understood. In the first and unproblematic sense, qualia are simply the ways things seem. In the second and problematic sense, qualia are understood to possess additional features; qualia are understood to be atomic, non-relational, ineffable, incomparable and incorrigibly accessible from the first-person point of view. The problem for Flanagan is not one snarled kite string, but rather two kite strings that have become ensnarled together:

One string is attached to an ordinary box kite that announces that there are certain ways things seem. The other string is attached to a gaudy kite that ostentatiously flies banners announcing that the ways things seem consists of a set of atomic, non-relational states that possess ineffable properties known deeply, incorrigibly, and exclusively from the first-person point of view. We can immediately improve matters by freeing the box kite from the extraneous one tangled up with it, the one that immodestly depicts qualia as possessing special problematic features...⁸⁴

Flanagan offers compelling reasons why it is important to retain the first notion of qualia. The concept of qualia is needed to distinguish mental states with “feel” from those without feel, to taxonomize various forms of subjective experience, and to set the agenda for hypothesis generation and theory construction at lower levels.⁸⁵

⁸⁴Flanagan, *Consciousness Reconsidered*, 62.

⁸⁵Ibid., 65-66.

Color experiences are typically appealed to as the quintessential examples of qualia. Opponents of naturalism have articulated some fairly robust claims about the nature of color qualia, arguing that subjective experience is beyond the predictive and explanatory powers of any physical theory. Dualistic conclusions are thus drawn about the metaphysical distinctness and irreducibility of qualia. These conclusions however are dubious.

Paul Churchland's Neurophilosophical Model of Qualia

In his article, "Chimerical Colors" Paul Churchland provides a striking example of work being done at the interface of psychology and neuroscience.⁸⁶ Color vision has long been a topic of psychological research with the psychologist Munsell mapping phenomenological color space nearly a century ago. More recently, cognitive neuroscientists have attempted to model and understand the mechanisms of human color vision. The Hurvich-Jameson network—what Churchland calls the "standard model of color perception"—is an attempt to explain the structure of color space in terms of the human visual system and has yielded some striking results.

What exactly is the H-J net? The H-J net is a theoretical model of human color vision that begins with retinal cone cells (inputs) and ends with color opponent cells (outputs). Cone cells are broadly tuned to different wavelengths of light: short (blue), medium (green), long (red), and background illumination. The three types of color-opponent cells (red-green, blue-yellow, and white-black) are the site of completion

⁸⁶Paul Churchland, "Chimerical Colors: Some Phenomenological Predictions from Cognitive Neuroscience," *Philosophical Psychology* 18, No. 5 (2005), in *Neurophilosophy at Work* (Cambridge, MA: MIT Press, 2007). The reader is encouraged to consult the colored figures that accompany Churchland's article.

between excitatory and inhibitory signals arriving from the cone cells. The H-J net converts a four-tuple of inputs: S, M, L, B into a three-tuple of outputs ($A_{B/Y}$, $A_{G/R}$, $A_{W/B}$).⁸⁷

These patterns of activation can be represented as coding vectors. The resting or default level is postulated to be 50% of the maximum activation levels. Thus the seen color gray results from a $\langle 50, 50, 50 \rangle$ coding vector. Color variations result when the excitatory/inhibitory signals coming in from the cones send the activation levels above or below 50. Increasingly saturated red results from excitatory signals from the cones, sending the number over 50, e.g. $\langle 50, 90, 50 \rangle$, while increasingly saturated green results from inhibitory signals sending activation levels below 50, e.g. $\langle 50, 10, 50 \rangle$.

Significantly, this configuration of coding vectors is structurally almost identical to the structure of phenomenal color space independently mapped by Munsell nearly a century ago. As Churchland notes:

[T]his peculiar configuration of possible *coding vectors* is structurally almost identical to the peculiar configuration, originally and independently reconstructed by Munsell, of possible *color experiences* in normal humans.... [T]he family of distance relations between all of the color experiences internal to the Munsell space is roughly identical with the family of distance relations between all of the coding triplets internal to the H-J spindle.... From precisely such global isomorphisms are speculative thoughts of intertheoretic identities likely to be born. The systematic parallels here described—though highly improbable on a priori grounds—become entirely nonmysterious if human color experiences (at a given point in one's visual field) simple *are* the output coding vectors (at a suitable place within some topographical brain-map of the retina) produced by some neuronal instantiation of the H-J net.⁸⁸

⁸⁷Ibid., 164-166.

⁸⁸Ibid., 169-70.

The H-J network also yields a number of explanations and predictions regarding colored after images. When one fixates on a red circle against a gray background and then moves one's gaze to a neutral gray background, a circular green afterimage is seen against the gray background. Why?

This happens because, when the (now fatigued) opponent cells representing the circular red stimulus are suddenly asked to fall back to representing a less-demanding middle-gray stimulus, they overshoot the required $\langle 50,50,50 \rangle$ coding vector by an amount equal to whatever fatigue or potentiation has been acquired in each of the three coding dimensions during the protracted exposure to the original red stimulus. That original red stimulus produced an initial coding vector of $\langle 50,95,50 \rangle$, but during protracted fixation, that initial vector slowly inches back to something like a vector or $\langle 50,55,50 \rangle$, thanks to the accumulated minus-40% fatigue in its middle $A_{G/R}$ element.

Accordingly, when the opponent cells in the fatigued area are suddenly asked to represent an objectively middle-gray stimulus, they can only manage to produce a vector of $\langle 50,10,50 \rangle$ —the coding triplet for an obvious middle green—instead of the $\langle 50,50,50 \rangle$ they would normally produce. For the $A_{G/R}$ cells in the affected circular area are, temporarily, too tired to respond normally. They produce a coding vector with a much-reduced middle component, an abnormal vector that represents green, not gray.⁸⁹

These predictions about colored after images are within the normal range of human color vision inside of the Munsell color spindle. But the H-J network predicts and explains the existence also of what Churchland calls “chimerical colors,” colors outside the Munsell color spindle but within the opponent cell activation cube.

But...what about all that *unused* space in the several upper and lower corners of the opponent cell activation cube? What would be the significance of a possible activation triplet *outside* the classical color spindle, a triplet somewhere in that fairly considerable volume of unused

⁸⁹Ibid., 174.

opponent-cell activation space.... [Y]ou might ask after the *phenomenological* significance of such an extraspindle activation vector.⁹⁰

Regarding such questions, the H-J network:

yields some novel and unappreciated predictions, and some novel and unappreciated explanations, concerning the qualitative characters of a considerable variety of color sensations possible for human experience, color sensations that normal people have almost certainly never had before, color sensations whose accurate descriptions in ordinary language appear semantically ill-formed or even self-contradictory.⁹¹

These chimerical colors are not the objective colors of real objects, but anomalous color representations within the cubical-cell activation space and can be produced by selective fatigue/potentialiation by prolonged fixation on a suitable color stimulus.⁹²

What are the philosophical implications of all of this? This account of color vision shows that the qualitative character of subjective experience is indeed *not* beyond the predictive or explanatory power of physical theory, and that there are good reasons to identify color qualia with neural coding vectors. The Munsell color solid neatly maps onto the H-J network's opponent cell activation cube, and as noted by Churchland, it is precisely such global isomorphisms that lead one to posit intertheoretic identities.

Here, as in those other cases from our scientific history, the principle intellectual motive for embracing the systematic color qualia/coding-vector identities proposed is simply the extent and quality of the predictive and explanatory unity that the relevant reduction provides.⁹³

⁹⁰Ibid., 180.

⁹¹Ibid., 161.

⁹²Ibid., 181.

⁹³Ibid., 191.

Further, the chimerical colors predicted and explained by the H-J net provide excess empirical content and thus another motive for embracing the identity of color qualia and coding vectors.

If these predictions are correct, they provide an *additional* motive for embracing the proposed reduction of color qualia to coding vectors. For it was no part of the motives—for the H-J net’s original proposal—that these particular experimental predictions be a part of the explanatory target. They were unanticipated, and they are faintly paradoxical on their face. They thus provide some “excess empirical content” beyond the original explanatory target, namely, our familiar experiences of the mundane colors of external objects.

Such excess empirical contents are familiar from the history of science. In the latter part of the nineteenth century, the assumption that light was identical to electromagnetic waves entailed that there should be such a thing as *invisible* light (an apparent contradiction, note well)... The parallel assumption, that human color representations or color qualia are identical with opponent cell-cell coding triplets in a neuronal instantiation of the H-J network, yields a similarly implausible prediction. There should exist color-qualia outside the qualitative range of the classical color spindle, qualia whose perfectly accurate descriptions violate our normal semantic expectations. The H-J theory further suggests how to produce such chimerical qualia—through opponent-cell fatigue/potential—so that we may test those unexpected predictions against our own experience.⁹⁴

Naturalistic Anti-Reductionism: Functionalism and the Autonomy of Psychology

Dualist arguments are the most far-reaching of anti-reductionist arguments in the philosophy of mind. Here the very possibility of vertical integration between phenomenal consciousness, scientific psychology, and neuroscience is ruled out, because consciousness cannot be explained in physical or functional terms because consciousness just isn’t physical. However, anti-reductionist arguments are also popular among

⁹⁴Ibid., 191-192.

physicalist philosophers of mind. These claims have important implications because they maintain that neuroscientific work will ultimately tell us nothing interesting about the mind.

Functionalism, Multiple Realizability, and the Autonomy of Psychology

Functionalists like Hilary Putnam and Jerry Fodor argued that mental states are properly characterized by their abstract causal role linking sensory inputs, mental states, and behavioral outputs, not by the “stuff” that *realizes* a particular mental state. Putnam and Fodor argued that different physical states could realize the same mental state: this is the doctrine of multiple realizability. Mental states are not type identical to specific brain states but rather token identical to any number of physical realizers. Functionalists saw the type identity envisioned in identity theory as overly ambitious. Identity theory, Putnam wrote:

...becomes still more ambitious when we realize that the brain-state theorist is not just saying that *pain* is a brain state; he is, of course, concerned to maintain that *every* psychological state is a brain state. Thus if we can find even one psychological predicate which can clearly be applied to both a mammal and an octopus (say “hungry”), but whose physical-chemical “correlate” is different in the two cases, the brain-state theory has collapsed. It seems to me overwhelmingly probable that we can do this.⁹⁵

Jerry Fodor appealed to multiple realizability arguments to show the failure of psycho-neural reduction and establish the autonomy of psychology from lower-level disciplines like neuroscience:

⁹⁵Hilary Putnam, "The Nature of Mental States," in *Mind and Cognition: A Reader*, ed. William Lycan (Oxford: Blackwell, 1990), 53.

If it turns out that the functional decomposition of the nervous system corresponds precisely to its neurological (anatomical, biochemical, physical) decomposition, then there are only epistemological reasons for studying the former instead of the latter. But suppose that there is no such correspondence? Suppose the functional organization of the nervous system cross-cuts its neurological organization. Then the existence of psychology depends not on the fact that neurons are so depressingly small, but rather on the fact that neurology does not posit the kinds that psychology requires.⁹⁶

Functionalist multiple realizability arguments thus cast doubt on the project of establishing reflective equilibrium between common sense psychology/phenomenal consciousness, scientific psychology, and neuroscience. There may be reflective equilibrium between common sense psychology and scientific psychology, but there will not be any sort of reflective equilibrium between scientific psychology and neuroscience. Productive research on mind will necessarily occur at levels higher than neuroscience. Neuroscience might well provide some interesting engineering details about how mind is implemented in *homo sapiens*, but neuroscience can no more explain the mind than physics can explain monetary exchange.⁹⁷ The moral of the functionalist account of multiple realizability is the autonomy of psychology from lower-level neuroscience.

Multiple Realizability Critiqued

Like the dualist arguments considered above, multiple realizability arguments seem to have a certain plausibility. However, when one dips into actual scientific practice, the claims of multiple realizability are counter-indicated. Scientific reductions and identities are contextual and domain-specific. When one fails to take into account

⁹⁶Jerry Fodor, "Special Sciences," in *The Philosophy of Science*, ed. Richard Boyd, Philip Gasper, and J.D. Trout (Cambridge, MA: MIT Press, 1991), 439.

⁹⁷Fodor explicitly makes this argument in "Special Sciences."

considerations of context and domain specificity, multiple realizability is in fact rampant throughout physical science. Consider, for example, the classic textbook identification of temperature with mean molecular kinetic energy. This claim, however, only applies to temperature *in a gas*.

Temperature in a solid, however, is identical to mean maximal kinetic energy, since the molecules of a solid are bound in lattice structures and hence restricted to a range of vibratory motions. Temperature in a plasma is something else entirely, since the molecular constituents of a plasma have been ripped apart. Even a vacuum can have a (“blackbody”) temperature, though it contains no molecular constituents. Temperature of classical thermodynamics is multiply realized microphysically in a variety of distinct physical states. Yet this is a “textbook” intertheoretic reduction and cross-theory identification. The reductions and identifications are specific to the domain of the physical state.⁹⁸

When one attends to such contextual factors, one sees, in fact, how little multiple realizability occurs at the interface of psychology and neuroscience. As John Bickle observes:

If radical multiple realizability really obtained among species in the actual world, contemporary neuroscientific experimental techniques built upon this assumption [of the continuity of underlying neural mechanisms within and across species] should bear little fruit. Why study the macaque visual system to investigate human visual processing, for example, if we can’t safely assume some continuity across species? Why should positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) reveal common areas of high metabolic activity during psychological task performance, both across and within individual humans—how down to a millimeter of spatial resolution? Standard neuroscientific experimental procedures and even clinical diagnostic tools would be hopelessly naïve in the face of significant multiple realizability. But these procedures do work (and are not hopelessly naïve).⁹⁹

⁹⁸*The Stanford Encyclopedia of Philosophy* (2006), s.v. “Multiple Realizability.”

⁹⁹Ibid.

William Bechtel and Jennifer Mundale have recently developed this “argument from neurosciences success” in some detail. Bechtel and Mundale bring a wealth of neuroscientific examples to bear on the issue and help pinpoint with some precision what goes wrong with multiple realizability arguments. Bechtel and Mundale consider neuroanatomical and neurophysiological approaches to brain mapping and also consider the role of neuroscience in the decomposition of the visual processing system.

Since the late 1800’s neuroscientists have been engaged in the project of mapping different areas of the brain. Korbinian Brodmann used anatomical tools to map the brain in the early part of the 20th century, resulting in the famous Brodmann maps of the brain still used today. Advances in cell staining techniques enabled Brodmann to recognize different types of neurons in the cortex and to discover that the cortex consisted of six different layers of neurons.¹⁰⁰ Significantly, Brodmann did his work *comparatively*. In defending his claim regarding six different layers in the cortex, Brodmann reports on preparations made from fifty-five different species.¹⁰¹ His goal in identifying different regions of the brain was due to the fact that he thought such anatomical differences would also be functionally significant.

Psychological criteria have also been central to brain mapping. Such psychological mapping is in fact older than neuroanatomical mapping, having its roots in the work of Franz Josef Gall. Here the research strategy focuses on deficits in psychological capacities that follow damage to particular parts of the brain. Language is

¹⁰⁰William Bechtel and Jennifer Mundale, "Multiple Realizability Revisited: Linking Cognitive and Neural States," *Philosophy of Science* 66 (1999): 179.

¹⁰¹*Ibid.*, 180.

perhaps the best known example, due to the pioneering work of Paul Broca and Karl Wernicke. Since this pioneering work lesion studies have become legion. But such lesion studies would not be successful if multiple realizability was true. As Bechtel and Mundale observe, “it is important to note that in interpreting these deficits, researchers implicitly reject multiple realization among human brains and assume that damage to a brain area in anyone will result in a deficit to a particular cognitive function that is performed by that area in undamaged brains.”¹⁰² Such work has also expanded to include stimulation studies and imaging studies.

Neuroscience has also proved successful in guiding the cognitive decomposition of visual processing in humans. As Bechtel and Mundale note, “if the taxonomies of brain states and psychological states were as independent of each other as the multiple realizability argument suggests, brain decomposition would be a poor guide for psychological decomposition;” however, understanding of psychological function is being fostered precisely by such an appeal to the brain and its organization.”¹⁰³

Bechtel and Mundale not only point to the *lack* of examples of multiple realizability at the psychology-neuroscience interface, they also pinpoint the intuitive appeal of multiple realizability arguments. The key issue involves contextuality. “Whenever one asks whether two items are the same or different, the question makes little sense unless one asks about sameness or difference with respect to some other

¹⁰²Ibid., 184.

¹⁰³Ibid., 190-191.

consideration.”¹⁰⁴ Particularly significant in terms of the psychology-neuroscience interface are issues concerning what Bechtel and Mundale call “grain size.” Multiple realizability arguments are flawed because they equivocate on the issue of grain size, differentiating psychological states in a coarse-grained manner while differentiating neurological states in a fine-grained manner. As Bechtel and Mundale write:

[O]ne diagnosis of what has made the multiple realizability claim as plausible as it has been is that researchers have employed different grains of analysis in identifying psychological states and brain states, using a coarse grain to identify psychological states and a fine grain to differentiate brain states. Having invoked different grains, it is relatively easy to make a case a case for multiple realization. But if the grain size is kept constant, then the claim that psychological states are in fact multiply realized looks far less plausible.¹⁰⁵

It is precisely this issue at stake in Putnam’s famous example of the multiple realizability of hunger in humans and octopi:

A human’s psychological state and that of an octopus might well be counted as the same insofar as they are associated with some general feature (such as food-seeking behavior in the case of hunger). But with respect to other considerations, a human psychological state may be considered different from that of an octopus, even if we limit the scope to mere behavior. Food-seeking behavior for the octopus is different from food-seeking behavior in the human if one is concerned about such things as how one seeks the food, what foods are sought, under what conditions, etc. This much seems simple and apparent, but the assertion that what we broadly call “hunger” is the same psychological state when instanced in human and octopi has apparently been widely and easily accepted without specifying the context for judging sameness.¹⁰⁶

¹⁰⁴Ibid.

¹⁰⁵Ibid., 203.

¹⁰⁶Ibid.

Vertical Integration: Beyond Reductive and Non-Reductive Physicalism

We have looked at several attempts to block the project of vertical integration between common sense psychology/ phenomenal consciousness, scientific psychology, and neuroscience. All attempts to deny that neuroscience *in principle* cannot offer an explanation of the conscious self has been denied. The project of vertical integration remains viable, but what *sort* of vertical integration is imaged? With anti-reductionist arguments blocked, does the alternative proposal for vertical integration have to be necessarily a reductive one? Are higher levels in the explanatory hierarchy destined to be lopped off by a matured neuroscience? This section answers no. Much of the debate between “emergentism and reductionism” or “reductive and non-reductive physicalism” rests upon a false dichotomy. As William Wimsatt observes:

An opposition between reduction and emergence forces people to take sides along an axis missing some of the most revealing cuts on the issue. One can be a reductionist and an emergentist too, with a proper understanding of these notions. Misunderstandings engender opposition to reductionism, and make emergence unnecessarily mysterious.¹⁰⁷

This section explores how one can move beyond the “reduction vs. autonomy” discussions that have dominated so much of contemporary philosophy of mind.

The Co-Evolution of Psychology and Neuroscience:

Reduction and Elimination

Because they reject the thesis of the principled autonomy of psychology from lower-level neuroscience, neurophilosophers and philosophers of neuroscience typically

¹⁰⁷William Wimsatt, "Emergence as Non-Aggregativity and the Biases of Reductionisms," in *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality* (Cambridge, MA: Harvard University Press, 2007), 274.

advocate a co-evolutionary research program between psychology and neuroscience; however, they differ sharply in their understanding of what co-evolution entails and what the outcome of that co-evolution is.¹⁰⁸ Robert McCauley has helpfully pointed out three ways in which co-evolution can be understood.¹⁰⁹ According to McCauley, psychology and neuroscience can be understood to co-evolve in a reductive (co-evolution_m), eliminative (co-evolution_s), or pluralist (co-evolution_p) manner.¹¹⁰

Co-evolution is frequently understood in either a reductive or eliminative manner. Advocates of both reduction and elimination understand the selection pressures in the co-evolution of psychology and neuroscience to be exerted almost exclusively in a bottom-up manner. The theories differ on the question of whether the reduction of psychology to neuroscience will be a smooth or bumpy one. The logical positivist notions of explanation and reduction used by early identity theorists envisioned scientific development in terms of smooth mapping of one higher-level scientific theory onto another lower level theory, as illustrated, for example, in the reduction of thermodynamics to statistical mechanics. This is primarily the understanding of co-evolution at work in co-evolution_m.

However, other reductions are not so smooth. Some reductions *eliminate* rather than smoothly *reduce* the phenomenon in question. Rather than a smooth one-to-one mapping, theory change in science often involves the displacement of one theory by the

¹⁰⁸For an early articulation of co-evolution see Patricia Churchland, *Neurophilosophy*.

¹⁰⁹Robert McCauley, "Explanatory Pluralism and the Co-Evolution of Theories in Science," in *Philosophy and the Neurosciences: A Reader*, ed. Pete Mandik William Bechtel, Jennifer Mundale, Robert Stufflebean (Oxford: Blackwell, 2001).

¹¹⁰*Ibid.*, 437-441.

other. Ptolemaic astronomy did not map neatly onto Copernican astronomy; rather the concepts of Ptolemaic astronomy were eliminated within Copernican theory. Phlogiston did not map neatly onto the theory of combustion; rather the concept “phlogiston” was eliminated by its successor theory. Looking at what they saw as the rather dismal explanatory success of common sense or folk psychology, the Churchlands argued that the mind-brain relationship was best understood in terms of elimination rather than reduction:

As the eliminative materialists see it, the one-to-one match-ups will not be found, and our common-sense psychological framework will not enjoy an intertheoretic reduction, *because our common-sense psychological framework is a false and radically misleading conception of the causes of human behavior and the nature of cognitive activity*. On this view, folk psychology is not just an incomplete representation of our inner natures; it is an outright *misrepresentation* of our internal states and activities. Consequently, we cannot expect a truly adequate neuroscientific account of our inner lives to provide theoretical categories that match up nicely with the categories of our common-sense framework. Accordingly, we must expect that the older framework will simply be eliminated, rather than be reduced, by a matured neuroscience.¹¹¹

Where the functionalists argued that psychology was autonomous from the neurosciences, the Churchlands’ position envisioned the elimination of common-sense and scientific psychology by a matured neuroscience. The eliminativist model understands the evolution of psychology and neuroscience based on the model of eliminative reductions that occur in scientific revolutions, hence co-evolution_s.

Beyond Autonomy and Reduction

Importantly McCauley suggests that neither co-evolution_m nor co-evolution_s are the most adequate accounts of the co-evolution that occurs at the interface between

¹¹¹Paul M. Churchland, *Matter and Consciousness*, 43.

psychology and neuroscience. In contrast to these accounts, McCauley defends a pluralistic account of co-evolution, $co\text{-}evolution_p$. McCauley's work is part of a broader naturalistic movement in the philosophy of science. Naturalistic philosophy of science attends to models at work in actual scientific practice, rather than imposing "one-size-fits-all" models on the whole of science. With this naturalistic turn, philosophy of science becomes the philosophy of a particular science, in this case psychology, biology, and neuroscience. Contemporary philosophers of biology, psychology, and neuroscience have called into question the adequacy of deductive-nomological models of explanation and reduction in these sciences, developing instead notions of mechanistic explanation and explanatory pluralism. Significantly, mechanistic and pluralistic models of explanation call into question the entire "autonomy vs. reduction" framework in which so much of the debate in the philosophy of mind has been cast. As William Bechtel observes:

There are other serious issues raised by the use of the reduction model as way of relating disciplines... Within this framework the focal questions have been whether or not psychological theories can be derived from neuroscientific ones. If so, psychological theories lose their autonomy. Accordingly, those arguing for the special status of psychology or other higher-level sciences have argued that such derivations are not possible... However... most neuroscience explanations do not take the form of D-N explanations in which phenomena are derived from laws, but rather are models of mechanisms. This casts a different light on the issue of reduction. Models of mechanisms are inherently reductionist: each proposed mechanism is designed to show how a phenomenon ascribed to a system is due to its constituent parts and their interaction. On the other hand, reduction no longer threatens the autonomy of the higher-level science: the higher level characterizes the interaction of processes, the lower level accounts for the performance of individual processes.¹¹²

¹¹²William Bechtel, "Philosophy Meets the Neurosciences," 17.

Wimsatt, Bechtel, and McCauley along with Peter Machamer, Lindley Darden and Carl Craver have developed accounts of mechanistic explanation that allow for substantial rapprochement between functionalism and identity theory without endorsing either autonomy or reduction. Mechanistic explanation is inherently multi-level and pluralistic. In an understanding of mechanistic explanation proposed by Carl Craver, an ideally complete mechanistic explanation of a system describes that system in three ways: as *isolated*, as *situated*, and as *constituted*.¹¹³ At the isolated (0-level), the system itself is described at its characteristic level. Situated (+1-level) and constituted (-1-level) descriptions operate above and below the isolated level. At the situated (+1-level) the system is situated in the context of some other systems. At the constituted (-1-level) the system is decomposed into its constitutive parts.¹¹⁴ As Thomas Polger notes, this model is neither inherently reductionist nor anti-reductionist:

The mechanistic model is not a “reductionist” style of explanation. In this respect it agrees with functionalism. Although mechanism permits multiple levels of causal-mechanical explanation, it is not “antireductionist” either, because that view still clings to the autonomy thesis. Classical “reductionist” inward- and downward-looking explanations emphasize the constitutive part of mechanistic integrations and neglect contextual explanation. In contrast, functional analysis emphasizes outward- and upward-looking contextual explanation and neglects constitutive explanation. Both, therefore, are incomplete from the point of view of mechanism.¹¹⁵

¹¹³Carl Craver, "Role Functions, Mechanisms, and Hierarchy," *Philosophy of Science* 68, no. 1 (2001); Peter Machamer, Lindley Darden, and Carl Craver, "Thinking About Mechanisms," *Philosophy of Science* 67 (March 2000). For a discussion of these points see also Polger, *Natural Minds*.

¹¹⁴Craver: 67-68; Polger, 198.

¹¹⁵Polger, 205.

Explanatory Pluralism and Co-Evolution

The explanatory pluralism inherent in mechanistic explanation suggests another way the co-evolution of psychology and neuroscience might be understood. This is a pluralistic understanding of co-evolution that McCauley calls co-evolution_p. Co-evolution_p is closely linked with the reductive account of co-evolution (co-evolution_m) but differs from it in important ways. Co-evolution_m understands the overwhelming majority of the selection pressures in the co-evolution of psychology and neuroscience to be exerted from the bottom up. While the upper-level theory may contribute in the process of discovery, providing an initial vocabulary and problems for research, sooner or later it must conform to the lower-level theory's explanation. In co-evolution_p by contrast,

theoretical proposals and the research they spawn at the higher level do not merely contribute to the process of discovery at the lower level. The upper level science provides a body of evidence against which the science at the lower level can evaluate competing models. This evidence is particularly useful, precisely because it frequently arises independently of the formulation of the specific lower-level models to whose assessment it contributes."¹¹⁶

McCauley is particularly concerned to mount a critique of the eliminative understanding of co-evolution, co-evolution_s. Here he draws upon an important distinction, initially made by Wimsatt, between *interlevel* and *intralevel* contexts.¹¹⁷ Eliminative reductions like those seen in scientific revolutions occur almost exclusively at an *intralevel* context rather than an *interlevel* context.

¹¹⁶McCauley, 448-449.

¹¹⁷See William Wimsatt, "Reductionism, Levels of Organization, and the Mind-Body Problem," in *Consciousness and the Brain*, ed. G. Maxwell G. Globus, and I. Savodnik (New York: Plenum Press, 1976).

The sorts of unequivocal eliminations of theories and ontologies that co-evolution countenances arise in *intralevel* contexts involving considerable incommensurability. These contexts concern changes *within* a particular science over time. They include the classic cases that philosophers group under the rubric of “scientific revolutions” – impetus, phlogiston, caloric fluid, and the like. Within a particular level of analysis, some newly proposed theory proves superior to its immediate predecessor with which it is substantially discontinuous. When the scientific community opts for this new theory, most traces of its predecessor rapidly disappear. Since they offer incompatible accounts of many of the same phenomena, the new theory *explains* the old theory *away*.¹¹⁸

However, the psychology-neuroscience interface is an *interlevel* context, and eliminations are exceedingly rare *between* levels.

Incommensurability in interlevel contexts neither requires the elimination of theories on principled grounds nor results in such eliminations in fact ...[T]he history of science and especially the history of late nineteenth- and twentieth century science offer no examples of large-scale interlevel theory elimination (particularly of the wholesale variety standard eliminativism and co-evolutions, envision) once the upper-level science achieves sufficient historical momentum to enjoy the accoutrements of other recognized sciences (such as characteristic research techniques and instruments, journals, university departments, professional societies, and funding agencies). The reason is simple enough. Mature sciences are largely defined by their theories and, more generally, by their research traditions; hence, elimination of an upper-level theory by a lower-level theory may risk the elimination of the upper-level scientific enterprise!¹¹⁹

This section has argued that one can embrace a neurophilosophical account of identity theory without eliminating higher-level sciences, but how exactly does identity theory function in the context of mechanistic explanation and explanatory pluralism?

¹¹⁸McCauley, 441.

¹¹⁹Ibid., 443-444.

Identity Theory and Explanation

A key question facing an identity theorist is the question of why one should posit identities rather than mere correlations. In dualist accounts, it is claimed that logical supervenience is necessary for reductive explanation and identification. Logical supervenience is determined on the basis of being able to conceptually analyze a phenomenon in physical and functional terms. Without such a necessary connection, the dualist critique claims that one can only assert correlations rather than identities. We have seen however, that there are ample reasons to question the explanatory criteria employed by dualists, but more needs to be said on the issue of “correlation’ vs. “identity.”

There is a great deal of ambiguity with the notion of “correlation” and a “neural correlate of consciousness.” This point is made well by Ilya Farber:

First off, it’s important to realize that there’s something misleading about the “correlate” part of “neural correlate of consciousness.” The term naturally focuses attention on correlational sources of evidence for NCC’s, such as the widely-cited experiments in which Nikos Logothetis found single cells whose activity covaried with monkeys’ behaviorally reported percepts. To treat the theories themselves as having this structure, however, is to mistake a noun for a verb. An NCC is *a correlate*, a *thing which corresponds* to consciousness; moment-to-moment psychophysical correlation is just one of many elements in this correspondence. All of the major NCC theories in fact draw multiple parallels, based on functional anatomy and pathology, on inter-species comparisons, and on interactions between the hypothesized NCC and other neural mechanisms which underlie related phenomena such as memory, dreaming, and emotion. The relation which these theories establish between the neural and phenomenal aspects of consciousness is thus something much richer and more complex than mere correlation; it is a type of *isomorphism*, a multidimensional mapping between entities, structures and dynamics in the twin domains of mind and brain...The goal of the NCC project is not to produce a causal model on which consciousness stands apart as a product of the brain, but rather to find the patterns of consciousness *within* the structure and dynamics of the brain. The methodology for pursuing this goal has already been charted out by researchers studying memory and perception: in

roughest outline, it involves functionally decomposing the cognitive process in question, functionally and physically decomposing the brain, and trying to find matching patterns amidst the bits on each side. This process is fundamentally *analogical* rather than correlational, and the relation that it attempts to establish is not one of causal interaction but one of *identity*.¹²⁰

Bechtel and McCauley have developed a scientifically informed heuristic account of identity theory much in line with Farber's comments.¹²¹ Bechtel and McCauley, in a manner very similar to Churchland, understand identity claims not so much as fixed metaphysical conclusions but as heuristics guiding scientific inquiry. Significantly, the notion of heuristic identity theory provides a powerful response to the correlation objection beloved by dualists. Hypothetical identities are not only conclusions of research; they also function as the premises of such research, as heuristics guiding scientific discovery. Essentially, heuristic identities guide scientific discovery through the converse of Leibniz's law—instead of the identity of indiscernables, emphasis is on the indiscernability of identicals in guiding subsequent research.¹²² What this means precisely is that what is learned about an entity or process under one description must apply to it under its other descriptions. Note that if the connection is a mere correlation, there would be no reason to expect such discoveries.

¹²⁰Ilya Farber, "How a Neural Correlate Can Function as an Explanation of Consciousness," *Journal of Consciousness Studies* 12, no. 4-5 (2005): 85.

¹²¹William Bechtel, "Mechanism and Phenomenal Experience: The Heuristic Identity Theory," (1993). Retrieved from: <http://mechanism.ucsd.edu/~bill/research/mercier/3rdlecture.pdf> (accessed 12/10/08); William Bechtel, and McCauley, Robert "Heuristic Identity Theory (or Back to the Future): The Mind-Body Problem against the Background of Research Strategies in Cognitive Neuroscience," *Proceedings of the 21st Annual Meeting of the Cognitive Science Society* (1999); William Bechtel, and Robert McCauley, "Explanatory Pluralism and the Heuristic Identity Theory," *Theory Psychology* 11, no. 6 (2001).

¹²²Bechtel, "Heuristic Identity Theory (or Back to the Future): The Mind-Body Problem against the Background of Research Strategies in Cognitive Neuroscience."

Consider, for example, genes and chromosomes. Cytologists in the early twentieth century identified chromosomes as paired structures in the cell nucleus, while the rediscovery of Mendel suggested paired factors accounting for heredity. As Bechtel notes, at the time there were only three similarities known to apply both to chromosomes and genes. There were other factors known just of genes or just of chromosomes. Hence, under heuristic identity theory the factors known to be true of genes were inferred to be true of chromosomes and vice versa.¹²³

For Bechtel, the importance of the heuristic perspective is the recognition that what gives credibility to the identity claim is the productivity of the claim in bringing two phenomena together. Identity theory does impose stringent demands—the application of Leibniz’s law. Applied to issues of consciousness and neuroscience, phenomenal experience guides discovery of neural mechanisms and neural mechanisms guide discovery about phenomenal experience.

The importance of the heuristic perspective is the recognition that what gives credibility to the identity claim is not... the correlations advanced at a given time, but the productivity of the claim in discovering new phenomenon that would only be expected on the basis of the identity. But that does not mean that the identity claim does not impose a tough demand. It does, but it is a demand not on evidence at a time but on the development of evidence in the course of research. What one should be seeking now is a promising identity claim between phenomenal experience and brain processes, one that points to future productive research. After advancing the identity claim, the demands imposed by Leibniz’s law apply. Future research must fill out the mapping by showing that everything known or learned about phenomenal experience maps onto neural processes, and vice versa. If the research is productive, then the

¹²³Bechtel, "Mechanism and Phenomenal Experience: The Heuristic Identity Theory."

identity claim will have proven its worth and the charge of mere correlation will lose plausibility.¹²⁴

Note that this is precisely the sort of process that occurs in Paul Churchland's "Chimerical Colors," where neuroscientific data is used to make predictions regarding color qualia. Churchland's remarks about the basis of identifying opponent-cell coding triplets and human visual color qualia echo the remarks of Farber, Bechtel, and McCauley:

The reader will note that, despite the nontrivial (but wholly defeasible) case laid out earlier, in support of the strict identity of human visual color qualia on the one hand and human opponent-cell coding triplets on the other; at no point did we establish, or even try to establish, that there is any sort of *necessary connection* between the two. I did not argue, nor claim, that the former are 'logically supervenient, upon the latter (cf. Chalmers 1996). I did not argue, nor do I believe, that the identity at issue is blessed by any form of metaphysical necessity (cf. Kripke 1972). Nor did I suggest that there is any form of 'lawlike' or 'nomological' connection between the two. As I have argued elsewhere, all of these diverse modal relations are philosophical extravagances or confusions imposed, post facto, on successful cases of historical intertheoretic reductions, all of which were achieved without the help of such modal relations, and none of which displays any one of them. Here, as in those other cases from our scientific history, the principal intellectual motive for embracing the systematic color-qualia/coding-vector identities proposed is simply the extent and quality of the predictive and explanatory unity that the relevant reduction provides.¹²⁵

As Ned Block and Robert Stalnaker point out, such "[i]dentities allow a transfer of explanatory and causal force not allowed by mere correlations."¹²⁶

¹²⁴Ibid.

¹²⁵Churchland, "Chimerical Colors," 191.

¹²⁶Ned Block and Robert Stalnaker, "Conceptual Analysis, Dualism, and the Explanatory Gap," *The Philosophical Review* 108, no. 1 (1999), 24.

Conclusion

The purpose of this chapter has been to argue that nothing stands in the way of bio-psychological explanation of the conscious self in terms of brain processes and that in fact psychoneural identity theories provide the most viable solution to the mind-body problem. Dualism and functionalism call into question the possibility of the vertical integration of phenomenal consciousness, scientific psychology, and neuroscience. This chapter has demonstrated that none of these arguments is compelling. However, it has also avoided equating identity theory with either reductive or eliminative forms of naturalism.

In light of the overall dissertation project, the point of the chapter was to establish that the identification of mind and brain is very well motivated scientifically and philosophically. Work of scientists and philosophers at the mind-brain, psychology-neuroscience interface stands in sharp contrast to the views of the religious dualists introduced at the beginning of this chapter. These issues will be revisited in Chapter Three, where attempts to integrate Christianity and Buddhism with the contemporary mind-sciences will be discussed. But what about the cross-cultural ubiquity of dualism? Surely that demands an explanation as well. This issue will form the topic of Chapter Two.

CHAPTER TWO
TOWARD A NATURALISTIC, BIO-PSYCHOLOGICAL
EXPLANATION OF DUALISM

Naturalism and the Ubiquity of Dualism

The previous chapter argued that dualism is no longer a compelling position in light of philosophical and scientific work at the mind-brain interface; however, the overwhelming majority of the world's population holds very robust dualist beliefs. People typically understand and experience themselves as separate and distinct from their bodies, believe in some sort of afterlife, and in "persons without bodies" (God, gods, spirits, ancestors, etc.). These beliefs are intimately linked with a variety of vivid dualist experiences in which the soul seems literally separate from the body.

How should one make sense of such dualist beliefs and experiences? If the arguments of Chapter One have been successful, significant doubt has been cast on the ability of dualism to withstand scientific and philosophical scrutiny. The view of the human person emerging from contemporary mind science and philosophy of mind differs radically from the view of the person entertained by most of humanity and by most of the world's religious traditions. If the arguments of Chapter One have been successful, we have every reason to believe that this view of the person is correct and reason to be skeptical about dualist claims about the human person, but if dualism is not true, the naturalist surely owes an explanation for the cross cultural ubiquity of dualism. This

chapter charts the course for a naturalistic explanation of dualism. Drawing upon the resources of neurophilosophy, neuroscience, and cognitive-developmental psychology, it argues that religious dualism is not *true*, but is “natural” in the sense that it is an outgrowth of our common sense understanding of our selves, other minds, and material objects.

Principled Humanist Objections to a Naturalistic, Bio-Psychological Explanation of Dualism

Chapter One addressed dualist and functionalist concerns about lower-level bio-psychological explanation of consciousness. However skepticism about lower-level bio-psychological explanations extends far beyond issues concerning consciousness in the philosophy of mind. Humanists, traditional social scientists, and scholars of religion typically favor “culture only” analyses and resist the claim that the sciences have much to contribute to the humanities.¹ The humanist emphasis on the autonomy of culture and on a methodology distinct from the natural sciences is itself rooted in a residual dualism. Of course, most scholars in the humanities explicitly reject mind-body dualism. Indeed, it would be hard to find a figure more reviled by humanist academics than Rene Descartes. Yet, at the same time, many humanists vigorously oppose any attempt to understand

¹In *What Science Offers the Humanities: Integrating Body and Culture*, Edward Slingerland cites the well-known writer Louis Menand as exemplifying such a stance. Menand writes of a crisis in the humanities, particularly literary studies, yet, he resists any attempt to find “consilience” between the sciences and the humanities. He is “certain that there is at least one thing that just *cannot* be wrong, that the sciences, especially the life sciences, have no place in the study of the human world” (Boyd , 2006, 19 as cited in Slingerland, 2008, 2).

human and social reality in biological terms.² An odd sort of false consciousness thus pervades large swathes of humanist academe: dualism is explicitly rejected while continuing to function operationally in some of the most basic assumptions of humanist and standard social scientific inquiry.

The crypto-dualist assumptions that pervade the humanities are perhaps most evident in Wilhelm Dilthey's principled distinction between the natural sciences (*Naturwissenschaften*) and the sciences of the free human spirit (*Geisteswissenschaften*). The German makes the principle of distinction between the sciences and the humanities explicit. The difference concerns the "objects" of inquiry: inert matter or the free human *Geist*. This radical difference in the objects of inquiry leads Dilthey to conclude that *Naturwissenschaften* and *Geisteswissenschaften* are also rooted in radically different epistemologies and methodologies. *Erklären* or reductive scientific explanation is the appropriate method of the *Naturwissenschaften*; however, the realm of the free human spirit can only be grasped by *Verstehen* or empathetic understanding.³

These dualist assumptions frequently go hand-in-hand with strong versions of social constructivism and arguments for the autonomy of culture. The link between dualism and social constructivism is perhaps most evident in the writing of the founder of modern sociology, Emile Durkheim:

Man is double. There are two beings in him: an individual being which has its foundation in the organism and the circle of whose activities is

²The examples of this "vigorous opposition" are legion. Steven Pinker provides a number of examples in his book, *The Blank Slate: The Modern Denial of Human Nature*. Opposition to the work of E.O. Wilson is probably the most notorious example.

³See Wilhelm Dilthey, *Introduction to the Human Sciences: An Attempt to Lay a Foundation for the Study of Society and History* trans. Ramon J. Batanzos (Detroit: Wayne State University Press, 1988).

therefore strictly limited, and a social being which represents the highest reality in the intellectual and moral order that we can know by observation – I mean society.⁴

Durkheim's dualist doctrine of the double man leads to a very specific understanding of social reality: society is seen as a superorganism inscribing the "blank slate" of the human mind. This understanding of social reality has very important methodological implications. To seek to explain social phenomena by appeal to psychology or biology is to miss the boat completely:

Every time that that a social phenomenon is directly explained by a psychological phenomenon, we may be sure that the explanation is false....The group thinks, feels, and acts quite differently from the way in which members would were they isolated.... If we begin with the individual in seeking to explain phenomena, we shall be able to understand nothing of what takes place in the group....Individual natures are merely the indeterminate material that the social factor molds and transforms. Their contribution consists exclusively in very general attitudes, in vague and consequently plastic predispositions.⁵

While Durkheim declared society off limits to lower level sciences, he saw religion as a social phenomenon and attempted to understand it accordingly; however, one very influential approach to religious studies rejects even Durkheim's sociological model as overly reductive. For scholars in the history and phenomenology of religion, the experience of the "sacred" is sui generis, and thus off limits to even higher level sciences like sociology. As Mircea Eliade writes:

A religious phenomenon will only be recognized as such if it is grasped at its own level, that is to say, if it is studied as something religious. To try to grasp the essence of such a phenomenon by means of physiology,

⁴Emile Durkheim, *The Elementary Forms of Religious Life*, trans., Joseph Swain (New York: Free Press, 1915/1966), 29.

⁵Emile Durkheim, *The Rules of the Sociological Method* (Glencoe, IL: Free Press, 1895/1962), 110.

sociology, economics, linguistics, art, or any other study is false; it misses the one unique and irreducible element in it – the element of the sacred.⁶

Hermeneutical distinctions between *Geisteswissenschaften* and *Naturwissenschaften*, coupled with social constructivist arguments about the autonomy of society have led not merely to a focus on interpretation over vertical explanation, but to principled exclusion of vertical explanation in the humanities and social sciences. Attempts at vertical integration are almost inevitably branded as reductionist. As Edward Slingerland writes:

The degree to which the mind versus body – and therefore understanding versus explanation – split has become entrenched in the modern university is reflected by the fact that, in the humanities, “reductionistic” has come to function as an immediately recognizable term of dismissive abuse: a claim that the understanding *Geist* has crossed the line and inappropriately slipped from *Verstehen* to *Erklären*, treating its subject as an object. People *do* seem fundamentally different to us than objects, which is why this understanding versus explanation distinction is able to gain a foothold in our minds. However, the conviction that the human can never be *explained* – that human-level phenomena can never be reduced to lower-level causal forces – takes this intuition a step further. The result is that the field of human inquiry has proudly wrapped itself in an impenetrable shell of *verstehen* and violently resists any attempt by the natural sciences to breach this boundary.⁷

The arguments and discussion in Chapter One provide an important point of departure in addressing these humanist concerns. The arguments of Chapter One give us good reason to be skeptical of dualist claims and thus of strong distinctions between *Naturwissenschaften* and *Geisteswissenschaften*. There simply is no *Geist* that exists independently of *Natur*. The arguments of Chapter One also called into question strong

⁶Mircea Eliade, *Patterns in Comparative Religion*, trans., Rosemary Sheed (Cleveland and New York: World, 1968), xiii.

⁷Edward Slingerland, *What Science Teaches the Humanities: The Integration of Body and Culture* (Cambridge: Cambridge University Press, 2008), 4.

versions of the “autonomy of psychology” thesis. This gives us cause for suspicion concerning other types of autonomy arguments. As Slingerland’s quote points out, autonomy arguments in the humanities and social sciences are frequently driven by a fear of reductionism. However, as Chapter One also pointed out, autonomy vs. reduction is a false dichotomy. The fear is that the natural sciences will somehow displace the humanities and even higher-level sciences. But such a notion of reductionism is either a bogeyman or, at best, highly antiquated. As Robert McCauley has pointed out:

The assumptions of antiquated conceptions of cross-scientific relations in terms of intertheoretic *reduction*, which envision the ability of lower level theories to displace higher level proposals and their ontologies, have haunted [humanists] from afar. Much recent work in the philosophy of science would aid considerably in exorcising these demons....Contrary to classical reductionism, displacing work at other levels is usually the last thing on scientists’ minds! Little, if any, evidence exists in twentieth century science of such displacements of theory or ontology arising as the result of investigations simultaneously carried on at multiple levels of analysis.⁸

That human-level interpretive strategies have much to offer is not being contested. What is being contested are claims of *principled autonomy* and *complete independence*. The discipline of religious studies originated in the West and in the context of Christian theology. The focus on particularities emphasized by many contemporary scholars in religious studies and the humanities has helped to correct many false assumptions about religion, but taken to extremes it creates its own problems. Here language might serve as an apt analogy. Language seems like the quintessential example of cultural diversity and plurality. But a purely cultural account of language leaves many questions unanswered.

⁸Robert McCauley, "Overcoming Barriers to a Cognitive Psychology of Religion," *Method & Theory in the Study of Religion* 12, no. 1 (2000): 149.

Why are the depth grammars underlying all languages so similar? How can one account for the ability of children's ability to become competent language users given the fact that they are exposed to it in a piecemeal manner? In the case of language this is all old history. Noam Chomsky and other cognitive linguists revolutionized the study of language in focusing on commonalities underlying the seemingly wild diversity of languages. The cognitive linguists were not calling into question the academic study of language and literature; they were asking a different kind of question. The burgeoning discipline of cognitive science of religion asks similar questions of religion. As cognitive anthropologist Scott Atran points out, there are many questions that have been left largely unanswered in the contemporary academic study of religion:

Why do *agent* concepts predominate in religion?
 Why are *supernatural-agent* concepts culturally universal?
 Why are *some* supernatural-agent concepts *inherently better* candidates for cultural selection than others?
 Why is it necessary, and how is it possible, to *validate* belief in supernatural agents that are logically and factually inscrutable?
 How is it possible to prevent people from deciding that the existing moral order is simply wrong or *arbitrary* and from *defecting* from the social consensus through denial, dismissal or deception?⁹

Atran's questions are closely linked to issues concerning dualism that this chapter is concerned to address, specifically the ubiquity of dualism in wildly diverse religious belief systems. The project is not so much to understand specific dualist beliefs, as to understand the "depth grammar" of dualism underlying religious particularities.

⁹Scott Atran, *In Gods We Trust: The Evolutionary Landscape of Religion* (Oxford; New York: Oxford University Press, 2002), 7.

Natural Sources of the Dualist Self: Toward a Bio-Psychological

Explanation of Religious Dualism

Principled objections to bio-psychological explanations of religion have been addressed, but many questions are still left unanswered. Even if one agrees that bio-psychological explanation of religion is possible, there is no agreed upon explanatory framework or anything remotely close to it. The most developed attempts at bio-psychological explanation of religion have come from neuroscience, cognitive-developmental psychology, and evolutionary biology, but there has been very little cross-fertilization between these approaches. All three of these approaches are relevant to the explanation of religious phenomena, but until very recently, issues concerning the conscious self and issues concerning dualism have not been a central concern in bio-psychological explanations of religion. Brain scans of religious experience are not the only or even the most important way the mind sciences can contribute to the bio-psychological explanation of religion. A bio-psychological account of the self and a bio-psychological account of dualism are equally important. In fact, discussions of scientific accounts of religious experience are apt to be mired in confusion if these foundational questions concerning the self and dualism are not addressed. The following sections offer a bio-psychological account of the natural sources of the dualist self. Its fundamental argument is that dualism emerges as a natural by-product of human beings' common sense understanding of their selves (folk phenomenology) and other selves (folk psychology).

Toward a Bio-Psychological Explanation of Religious Dualism:

Insights from Neuroscience and Neurophilosophy

A central premise of this dissertation is that souls do not exist, nor do strongly emergent selves. We can be realists of a sort about phenomenal experience and the phenomenal self, but there is nothing mysterious about this. The self is not something independent of the brain, or something miraculously squirted out by the brain. The self is merely a vastly complex representational structure within the human brain and experienced “from the inside” as it were. In this sense, Francis Crick is absolutely right that we are “nothing but a pack of neurons,”¹⁰ and this hypothesis is indeed astonishing when seen in light of dualist and emergentist theories of mind that understand consciousness as a further fact or an added extra.

This section explores how the conscious self can be explained in a vertically integrated manner and uses this information to shed light on the ubiquity of dualism. This bio-psychological account of the self bolsters the arguments in favor of the naturalistic, vertical explanation of the conscious self offered in Chapter One. In that chapter, it was argued that principled objections to a naturalistic account of the conscious self were not compelling, and argued that the hard problem of consciousness is best seen as a real but empirically tractable problem. Chapter One showed *that* the hard problem was empirically tractable; this section shows *how* the hard problem is empirically tractable. Actual scientific, bio-psychological research programs and theories of consciousness have been, and continue to be, developed (and very well funded) despite the nay-saying

¹⁰Francis Crick, *The Astonishing Hypothesis: The Scientific Search for the Soul* (New York: Simon & Schuster, 1994), 3.

arguments of dualists, and naturalism becomes compelling not so much as a refutation of dualism but because of its constructive potential in generating testable, progressive research programs and because of the explanatory success of such programs. The proof of naturalism is in the pudding, so to speak. However, a bio-psychological account puts one in a position not only to defend naturalism but to explain dualism. As such, the implications of such an account extend far beyond neuroscience, cognitive science, and the philosophy of mind. Because of the ubiquity of immaterial mind concepts (i.e., “persons without bodies”) within religious traditions, such an account is surely germane to theology, philosophy of religion, and religious studies more generally.

While scientific accounts of consciousness have exploded in recent years, this section focuses particularly on the work of Antonio Damasio and Thomas Metzinger. Damasio and Metzinger’s work is significant for a number of reasons. First, both thinkers take the hard problem of consciousness seriously, while seeking an empirical solution to it. Second, both thinkers are concerned with the issue of the *conscious self* and not merely with the issue of consciousness. Third, the work of these thinkers shares a number of important and striking similarities, particularly the emphasis on the conscious self as a complex representational phenomenon. Finally, Metzinger’s work sheds a great deal of light on the ubiquity of dualism and the explanation of that ubiquity. This section first explores how Damasio and Metzinger understand and pose the problem of the conscious self. It then turns to Damasio’s representationalist account of the origins and evolution of the self and to Thomas Metzinger’s representational account of the conscious self as a

transparent phenomenal self-model (PSM). Finally, it is shown how these representational accounts of the self help to explain the ubiquity of dualism.

Defining the Problem of the Conscious Self

Two temptations are almost impossible to avoid in terms of the hard problem: making the hard problem an impossible mystery or turning the hard problem into an easy problem. Dualists and mysterians are guilty of the first mistake; deflationist materialists are guilty of the second. Explaining the conscious self is not intractable, but its difficulty should not be underestimated either. Antonio Damasio and Thomas Metzinger are both thinkers who take the hard problem seriously while aiming for an empirical solution to it. The hard problem is the problem of how objective brains give rise to subjective experience. The hard problem is typically posed as a problem concerning consciousness, but it is important to note that the problem of the *conscious self* is implied in the definition of the hard problem. The problem is not just how objective brains give rise to experience but how they give rise to *subjective* experience. Consciousness and the conscious self are thus both parts of the hard problem. Metzinger captures these important distinctions well:

What is it that makes consciousness such a special target phenomenon? In conscious experience *a reality is present*. But what does it mean to say that, for all beings enjoying conscious experience, necessarily *a world appears*? It means at least three different things: In conscious experience there is a world, there is a self, and there is a relation between both—because in an interesting sense this world appears *to* the experiencing self. We therefore distinguish three different aspects of our original question. The first set of questions is about what it means that a reality *appears*. The second set is about how it can be that this reality appears to *someone*, to a subject of experience. The third set is about how this subject becomes the center of its own world, how it transforms the appearance of a reality

into a truly *subjective* phenomenon by tying it to an individual first-person perspective.¹¹

As Damasio more colloquially puts it, there are thus at least two problems implied in the hard problem: “the problem of how the movie-in-the-brain is generated, and the problem of how the brain also generates the sense that there is an owner and observer for that movie.”¹²

An Evolutionary Account of the Conscious Self: Antonio Damasio

One source of mischief with the mind-body problem is that in philosophical discussions it is almost always approached in a synchronic manner, i.e., how do objective brains create subjective experiences *right now*? The diachronic question of how something like a conscious self could emerge in an evolutionary context is seldom addressed, but viewing the conscious self in a diachronic, evolutionary manner goes a long way in making the synchronic problem much less mysterious. It is significant to note that viewing biological phenomena in an exclusively synchronic manner creates similar befuddlement. Viewed synchronically it is nearly impossible to see how anything other than *ex nihilo* intelligent design could generate the complex design of organisms. Diachronic, gradational thinking introduced in nineteenth century geology and biology culminating in the work of Darwin changed all of this. Explaining consciousness and explaining the apparent design of organisms are not the same thing, but there are significant analogies, and a diachronic approach to the mind-body problem may prove as

¹¹Thomas Metzinger, *Being No One : The Self-Model Theory of Subjectivity* (Cambridge, MA: MIT Press, 2003), 15.

¹²Antonio R. Damasio, *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*, 1st ed. (New York: Harcourt Brace, 1999), 11.

fruitful for the explanation of consciousness as it was for the explanation of the apparent intelligent design of organisms.

One thing a diachronic, evolutionary approach draws attention to is the centrality of body and environment/world in thinking about the conscious self. The hard problem cannot be solved by approaching it exclusively in a synchronic manner, and it also cannot be solved by thinking of mind and brain in an isolated and atomistic manner. It is easy to think of the mind performing abstract operations that seem disconnected with the life of an organism, and it is possible to view the brain in an atomistic “brain-in-the-vat” fashion apart from the needs to the rest of the body, but it is precisely the mind-brain’s connection with the entire organism and with the external environment that explains its evolutionary origins and helps to dissolve the hard problem. When the mind-brain problem is approached in a diachronic, evolutionary manner, as the mind-brain-body-world problem, one begins not with minds and brains but rather with simple single-celled organisms (bodies) in an external environment (world). Damasio’s work draws particular attention to the long natural history of the self. Even at the very rudimentary level of single-celled organisms, the boundary that defines the organism marks off its internal environment (“the internal milieu”) from an external environment.

Life is carried out inside a boundary that defines a body. Life and the life urge exist inside a boundary, the selectively permeable wall that separates the internal environment with the external environment. The idea of organism revolves around the existence of that boundary.... Life needs a boundary. I believe that minds and consciousness, when they eventually appeared in evolution, were first and foremost about life and the life urge within a boundary.¹³

¹³Ibid., 137.

It was the French biologist Claude Bernard who first noted an organism's need to maintain a consistent "internal milieu" in an environment that is in constant flux. The British biologist W.B. Cannon amplified Bernard's work, coining the term *homeostasis*: "the coordinated physiological reactions which maintain most of the steady states of the body...and which are so peculiar to the living organism."¹⁴ Even the simplest single-celled organisms possess an "internal milieu" that must be kept constant in light of a constantly changing environment. It is precisely here that rudiments of the mind-brain can thus be found. Damasio argues that the "internal milieu" of an organism is thus the precursor to more robust notions of self:

The specifications for survival that I am describing here include: a boundary; an internal structure; a dispositional arrangement for the regulation of internal states that subsumes a mandate to maintain life; a narrow range of variability of internal states so that those states are relatively stable. Now consider these specifications. Am I describing *just* a list of specifications for the survival of a simple living organism, or could it be that I am also describing some of the biological antecedents of the sense of self – the sense of a single, bounded, living organism bent on maintaining stability to maintain its life? I would say that I might be describing either. It is intriguing to think that the constancy of the internal milieu is essential to maintain life *and* that it might be a blueprint and anchor for what will eventually become a self in the mind.¹⁵

Nervous systems greatly enhance these capacities for coordinating the inside of an organism with its external environment. This first occurs entirely unconsciously through the function of what Damasio calls the proto-self:

I have come to conclude that the organism, as represented inside its own brain, is a likely biological forerunner for what eventually becomes the elusive sense of self. The deep roots for the self, including the elaborate

¹⁴Ibid., 138.

¹⁵Ibid., 136.

self which encompasses identity and personhood, are to be found in the ensemble of brain devices which continuously and *nonconsciously* maintain the body state within the narrow range and relative stability required for survival. These devices continually represent, *nonconsciously*, the state of the living body, along its many dimensions. I call the state of activity within the ensemble of such devices the *proto-self*, the nonconscious forerunner for the levels of self which appear in our minds as the conscious protagonists of consciousness: core self and autobiographical self.¹⁶

What becomes clear from this discussion of internal milieu and proto-self is that the organism needs to “know” its own needs, be aware of threats to its internal milieu, and engage in coordinated movement in response to an external environment that is continually changing. As Damasio puts it, “body-minded minds help save the body.”¹⁷ “Survival depends on finding and incorporating sources of energy and on preventing all sorts of situations which threaten the integrity of living tissue.”¹⁸ This requires a complex coordination of action and image. A device that facilitates this complex coordination would have conferred enormous advantage to an organism possessing it:

If actions are at the root of survival and if their power is tied to the availability of guiding images, it follows that a device capable of maximizing the effective manipulation of images in the service of the interests of a particular organism would have given enormous advantages to the organisms that possessed the device and would probably have prevailed in evolution. Consciousness is precisely such a device. The pathbreaking novelty provided by consciousness was the possibility of connecting the inner sanctum of life regulation with the processing of images.¹⁹

¹⁶Ibid., 22.

¹⁷Ibid., 143.

¹⁸Ibid., 23.

¹⁹Ibid., 24.

How exactly does consciousness and the conscious self emerge from the nonconscious proto-self? According to Damasio, consciousness/ the conscious self emerges as the organism is affected by an object. Importantly, for Damasio, this is necessarily a second-order activity. The state of the organism is constantly being mapped by the proto-self; any object the organism interacts with is also mapped within the sensory and motor structures of the brain.²⁰ These sensorimotor representations cause changes to the representations pertaining to the organism, i.e., the proto-self. Consciousness and the conscious self emerge as these changes are re-represented in higher-order maps representing the relationship between organism and object.²¹ Consciousness is a form of higher-order representation: “The first basis for the conscious you is a feeling which arises in the re-representation of the nonconscious proto-self in the process of being modified.”²² This form of consciousness is what Damasio calls “core consciousness”: “a simple, biological phenomenon, it has one single level of organization; it is stable across the lifetime of the organism; it is not exclusively human; and it is not dependent on conventional memory, working memory, reasoning, or language.”²³ Importantly, a core self emerges in this second-order process of organism being affected by an object:

Looking back, with the license of metaphor, one might say that the swift, second-order nonverbal account narrates a story: that of the organism

²⁰Ibid., 169.

²¹Ibid.

²²Ibid., 172.

²³Ibid., 16.

caught in the act of representing its own changing state as it goes about representing something else. But the astonishing fact is that the knowable entity of the catcher has just been created in the narrative of the catching process.²⁴

The sense of self that emerges from core consciousness Damasio calls the core self, “a transient entity, ceaselessly recreated for each and every object with which the brain interacts.”²⁵ This form of consciousness and conscious selfhood is created in pulses, in the constant interaction of organism and object. Core consciousness and the core self emerge in the process of the proto-self being modified by interaction with the external environment. Core consciousness and the core self in turn generate extended consciousness and the autobiographical self. This is a robust sense of self; the self Daniel Dennett has called “the center of narrative gravity.”²⁶ Most simply, it is an “an organized record of past experiences of an individual organism.”²⁷

The emergence of consciousness and the conscious self is of clear evolutionary significance. As Damasio understands it, consciousness introduces a new means of maintaining the internal milieu of an organism and thus preserving homeostasis.²⁸ It does this not by replacing the functions of the proto-self that can be performed with great efficiency in an unconscious manner, but by helping the organism respond more efficiently to changes in the external environment.

²⁴Ibid., 170.

²⁵Ibid., 17.

²⁶Daniel C. Dennett, *Consciousness Explained* (Boston: Little, Brown, 1991), 412-430.

²⁷Damasio.

²⁸Ibid., 303.

Creatures with consciousness have some advantages over those that do not have consciousness. They can establish a link between the world of automatic regulation (the world of basic homeostasis that is interwoven with the proto-self) and the world of imagination (the world in which images of different modalities can be combined to produce novel images of situations that have not yet happened). The world of imaginary creations—the world of planning, the world of formulation of scenarios and prediction of outcomes—is linked to the world of the proto-self. The sense of self links forethought, on the one hand, to preexisting automation, on the other.

Consciousness is not the sole means of generating adequate responses to an environment and thus achieving homeostasis. Consciousness is just the latest and most sophisticated means of doing so, and it performs its function by making way for the creation of novel responses in the sort of environment which an organism has not been designed to match, in terms of automated responses.²⁹

Damasio helps us see how a complex phenomenon like the conscious self is related to more fundamental issues concerning the relationship between organisms in particular environments. Such an approach goes a long way toward demystifying the conscious self. It also becomes clear that the notion of representation does a great deal of conceptual work in Damasio's account.

Natural Sources of Religious Dualism I: Selves as Transparent

Phenomenal Self Models: Thomas Metzinger

Metzinger also offers a representationalist analysis of consciousness with many of the same emphases as Damasio. In Metzinger's analysis, the self emerges as a phenomenal self model (PSM) is placed inside of a phenomenal world model:

First, our brains generate a world-simulation, so perfect that we do not recognize it as an image in our minds. Then, they generate an inner image of ourselves as a whole. This image includes not only our body and our psychological states but also our relationship to the past and the future, as well as to other conscious beings. The internal image of the person-as-a

²⁹Ibid., 304.

whole is the phenomenal Ego, the “I” or “self” as it appears in conscious experience.... The phenomenal Ego is not some mysterious thing or little man inside the head but the content of an inner image—namely, the conscious self-model, or PSM. By placing the self-model within the world model, a center is created. It is the origins of what philosophers often call the first-person perspective.³⁰

Damasio and Metzinger both highlight how something like a first-person perspective might emerge. The first-person perspective is a crucial aspect of consciousness phenomenology, but it is not the only one. Our conscious experience does not seem to have a character of a representational model and certainly does not seem very brainy: on the contrary, we seem to be outside of our brains and in direct contact with reality. Metzinger, in particular, draws attention to this aspect of our conscious experience, what he calls the *transparency* of phenomenal self model (PSM). The notion of the transparency that Metzinger is drawing upon was originally introduced by G.E. Moore:

... the fact that when we refer to introspection and try to discover what the sensation of blue is, it is very easy to suppose that we have before us only single term. The term “blue” is easy enough to distinguish, but the other element which I have called “consciousness”—that which a sensation of blue has in common with a sensation of green—is extremely difficult to fix... And in general, that which makes the sensation of blue a mental fact seems to escape us; it seems, if I may use a metaphor, to be transparent—we look through it and see nothing but the blue; we may be convinced that there *is something*, but *what* it is no philosopher, I think, has yet clearly recognized.³¹

³⁰Thomas Metzinger, *The Ego Tunnel : The Science of the Mind and the Myth of the Self* (New York: Basic Books, 2009), 7.

³¹G.E. Moore, “The Refutation of Idealism,” *Mind* 12 (1903), 446 as quoted in Metzinger, *Being No One : The Self-Model Theory of Subjectivity*, 163.

When something is transparent, one does not see it but sees *through* it, “we do not see the window but only the bird flying by;” we are not aware of the medium through which information reaches us.³² It is precisely transparency—the fact that we never see the self model but only see through it—that turns a system model into a self-model. “[T]he transparency of representational structures is the decisive criterion for turning a model into an appearance, into an apparent reality.”³³ In Metzinger’s self-model theory of subjectivity, the self is understood as a representational model, a phenomenal self-model (PSM), that is not recognized *as a model*.

Nobody ever *was* or *had* a self. All that ever existed were conscious self-models that could not be recognized *as* models. The phenomenal self is not a thing, but a process—and the subjective experience of *being someone* emerges if a conscious information-processing system operates under a transparent self-model....[T]he conscious self-model of human beings is the best invention Mother Nature has made. It is a wonderfully efficient two-way window that allows an organism to conceive of itself *as a whole*, and thereby causally interact with its inner and outer environment in an entirely new, integrated, and intelligent manner. Consciousness, the phenomenal self, and the first-person perspective are fascinating *representational* phenomena that have a long evolutionary history, a history which eventually led to the formation of complex societies and a cultural embedding of conscious experience itself.³⁴

The phenomenology of transparency is phenomenology of immediacy, of apparently direct perception, due to the fact that the representational character of the contents of conscious experience is itself not accessible to conscious experience.³⁵

“Inaccessible to conscious experience is the simple fact that this is taking place within a

³²Metzinger, *Being No One: The Self Model Theory of Subjectivity*, 170.

³³Ibid., 169.

³⁴Ibid., 1.

³⁵Ibid., 169.

medium.”³⁶ The transparency of the conscious self makes perfect sense in an evolutionary perspective. The transparency of representation allows for what *seems to be* immediate contact with reality, the sense that “I am present in a world outside my brain and in immediate contact with reality.” For instance, I see an animal. “I” and “the animal” are complex representational structures in my brain, but we do not need to know this. It does not help us in “flagging the dangerous present”³⁷ at all. In fact, it would spectacularly hinder our ability to do so. It is easy to see the massive selective advantages transparency would allow. Transparency allows for efficient action in the present (e.g., escaping an attacking animal). Note too, that the baseline “zero-world” model provided by the transparency of consciousness becomes essential as human beings develop more sophisticated off-line meta-representational capacities such as conscious planning, thinking, and deliberating. Such conscious deliberation and planning are only intelligible against a baseline reality provided by the transparent self model. Transparency allows for the introduction of a reality/representation, appearance/reality distinction.

The phenomenology of transparent representation gives us the sense of being outside our brains and in immediate contact with the world. It is important to realize that what makes good sense from a pragmatic evolutionary standpoint does not necessarily make sense from a scientific-philosophical point of view. It’s difficult to estimate the number of errors that result from a failure to recognize this. It’s an easy slide from a pragmatic-evolutionary folk phenomenology to folk metaphysics. There, the transparency

³⁶Ibid., 169-170.

³⁷The psychologist Richard Gregory uses this term to indicate what he considers to be one of the major evolutionary functions of consciousness.

of consciousness and the conscious self makes us naïve realists about the external world and dualists about the self (dualism might be thought of as naïve realism about the internal world). We conclude that we must *really* be outside of our brains and in immediate contact with the world. But understanding ourselves as literally outside of our brains and in immediate contact with the world makes no sense scientifically. How can the self interact with the world without violating conservation laws? How can a non-physical self emerge in the process of evolution? Here it is clear that phenomenology is *not* ontology, we are not really outside our brains, and we are not really in immediate contact with reality—dualism and naïve realism are elaborate cognitive illusions, virtual realities. The Finnish neuroscientist Annti Revonsuo puts this well:

[M]other nature...clothed us, that is, the phenomenological level of organization, with a built-in “out-of-brain” experience. But, remember, we are not really out of our brains in our experiences—how could we if all the required neural machinery sits tightly there.... No, we are only virtually out of the brain and in the external world. In fact, we did not invent VR [Virtual Reality] at all—evolution invented it for us millions of years ago. We have merely invented one fresh way to use, with the help of computer technology, the natural VR machine in our brain.³⁸

The virtual reality metaphor is a great heuristic tool for thinking about the hard problem in an empirically tractable manner, but it also has limitations. In artificial virtual reality, there is still a conscious self experiencing the virtual reality, but Metzinger and Revonsuo are not advocating a homuncular metaphor of consciousness: the conscious self is itself part of the virtual reality. Metzinger uses the image of a “total flight simulator” to illustrate this point:

³⁸Annti Revonsuo, "Consciousness, Dreams, and Virtual Realities," *Philosophical Psychology* 8 (1995): 14.

The brain is like a *total flight simulator*, a self-modeling airplane that, rather than being flown by a pilot, generates a complex internal image of itself within its own internal flight simulator. The image is transparent and thus cannot be recognized as an image by the system. Operating under the condition of a naïve-realistic self-misunderstanding, the system interprets the control element in this image as a nonphysical object: the “pilot” is born into a virtual reality with no opportunity to discover this fact.... If the virtual self functions extremely well, the organism using it is completely unaware of its “as if” nature. The self-model activated in the human brain has been optimized over millions of years. The process that constructs it is fast, reliable, and has a much higher resolution than any of today’s virtual-reality games. As a result, the virtuality of the phenomenal self-model tends to be invisible to the user. But strictly speaking, it is simply the best hypothesis the system has about its own current state—presented in a new, highly integrated data format.³⁹

The transparency of the phenomenal self model and the virtual reality metaphor has very significant implications. They provide a way to take phenomenology seriously while at the same avoiding dualist intuitions, both the Scylla of deflationist materialism and the Charybdis of dualism are avoided. The notion of the transparent phenomenal self model allows one to view the self as part of a vertical explanatory hierarchy. As Metzinger notes:

In introducing the working concept of a PSM I claim that it constitutes a *distinct* theoretical entity. That is, I claim that it is not only something that can meaningfully be described on a number of different levels of description mirroring each other in a heuristically fruitful manner but that it is something that can be *found* by suitable empirical research program. And it can be found on *every* level of description.⁴⁰

How can we understand this? What is commonly referred to as “the self” is identical to a phenomenal self-model (PSM). As such, the PSM, can be viewed from a first person and third person perspective and at various explanatory levels found in scientific psychology

³⁹Metzinger, *The Ego Tunnel : The Science of the Mind and the Myth of the Self*, 108.

⁴⁰Metzinger, *Being No One : The Self-Model Theory of Subjectivity*, 303.

and neuroscience (i.e., representational, functional, neurological). When viewed from a first person phenomenological perspective, the self model is not experienced as a model due to the transparent nature of the self model. It is thus a great temptation to view the phenomenological level as entirely apart from this explanatory hierarchy (as in robust versions of dualism) or as an additional, autonomous vertical level in the explanatory hierarchy, as something over-and-above the descriptive levels found in scientific psychology and neuroscience (as in property dualism and emergentism). The phenomenological level is not something “over-and-above” the representational level. The phenomenal self *is* the content of the transparent PSM and the transparent PSM can be identified with some global neural correlate of consciousness. The first person-third person shift is a horizontal relation and not a vertical one: subjective physical facts and objective physical facts are the same thing viewed from different perspectives, from “the inside” and “the outside” as it were.

But the significance of the notion of a transparent self-model extends beyond the scientific and philosophical study of consciousness. The phenomenal self-model is not recognized *as a model*. As a result all of our experience seems to be “out-of-brain” experience, thus providing a potent source of dualist intuitions. However, the brain is capable of generating an even more robust sense of dualism. The brain can generate the sense of not only being outside our brain but being outside our body. Research on how the brain represents the body not only provides empirical support for the notion of the self as a virtual, representational model, it also helps to explain another very important

source of dualist intuitions—out-of-body experiences (OBEs). Research on bodily representation and specifically on OBEs forms the topic of the next section.

Natural Sources of Religious Dualism II:

Thomas Metzinger and Olaf Blanke on Out-of-Body-Experiences

OBEs are a well-reported cross-cultural phenomenon and occur in about 10 percent of the general population. For those undergoing these vivid experiences dualist intuitions are almost inevitable. As Metzinger writes:

For anyone who actually has had that type of experiences it is almost impossible not to become an ontological dualist afterwards. In all their realism, cognitive clarity, and general coherence, these phenomenal experiences almost inevitably lead the experiencing subject to conclude that conscious experience can, as a matter of fact, take place independently of the brain and the body: what is phenomenologically possible in such a clear and vivid manner must also be metaphysically possible or actually the case.⁴¹

Much neuroscientific light has been shed on issues of bodily self-representation in recent years, and, as Metzinger and the Swiss neuroscientist Olaf Blanke have shown, progress can be made by viewing OBEs in such a context.

The previous section was at pains to point out that the self is not a “thing” but a pragmatic model—the brain’s best guess about the organism’s status in relation to the present external environment. This pragmatic, virtual character of the self has been documented in a number of experiments on bodily representation. Representing one’s body is no small thing. The sense of one’s body that one has at every instant requires the interpretation of inputs from various sensory modalities in order to integrate and bind all

⁴¹Thomas Metzinger, “Out-of-Body Experiences as the Origin of the Concept of A “Soul”,” *Mind & Matter* 3, no. 1 (2005): 78.

of this information together. Understanding how everything “comes together” to create the apparent cohesiveness and unity of conscious experience is one of the most significant problems in neuroscience, the so-called “binding problem.” The binding problem is not, as yet, completely understood, but the failure to bind information and the various multi-sensory conflicts and illusions generated by the failure of sensory integration in the brain are increasingly subject to empirical investigation. As Metzinger and Blanke have shown, far from being mystical, occult phenomena, OBEs can be fruitfully viewed in such a context.

In clinical literature, OBEs are classified as a type of autoscopic illusion.⁴² Autoscopic illusions all involve seeing a duplicate of one’s own body. Metzinger has helpfully placed discussion of these complex whole-body illusions alongside body *part* illusions. Vivid examples of the pragmatic nature of self-modeling and bodily representation – i.e., its virtual character – can be seen in research on body part attribution and location. In an experiment by University of Pittsburgh psychiatrists, Matthew Botvinick and Jonathan Cohen, healthy experimental subjects experienced an artificial rubber hand as their own hand; this has come to be known as the rubber hand illusion (RHI). In this experiment, one’s own hand is hidden from view and one observes a fake rubber hand in front of them. Both the hidden real hand and the seen fake hand are then stroked with a probe. Because visual information typically overrides other sensory modalities, observation of the stroking of the fake hand causes people to attribute the fake

⁴²Christine Mohr and Olaf Blanke, “The Demystification of Autoscopic Phenomena: Experimental Propositions,” *Current Psychiatry Reports* 7, no. 7 (2005).

hand to their own bodies (“to feel like it is my hand”). Sensations of having a full blown “virtual” arm—a connection from the shoulder to the fake hand have also been reported in other RHI experiments, along with a “proprioceptive drift”—a mislocalization of one’s hand toward the fake hand.⁴³

Phantom limbs are another closely related phenomenon. In such cases, patients have the “persistent and unmistakable impression” that a lost limb is still part of their body. In certain cases, the phantom limb is paralyzed, creating the impression that the absent limb cannot be moved. Working with a patient, Philip, with a paralyzed phantom limb, V.S. Ramachandran and his colleagues provided another striking example of the dynamic and pragmatic nature of self-modeling, underscoring the degree to which the self-model depends on perceptual and contextual information.⁴⁴ Ramachandran created a “virtual reality box” by placing a mirror inside a cardboard box open at the top with two holes cut in the front of the box to either side of the mirror. Philip was instructed to place his real arm and his paralyzed phantom limb into the box. He was then told to observe the reflection of his real hand in the mirror. The mirror image of his right hand was used to create the visual illusion that he actually did have two hands. Next, he was asked to make symmetrical movements with both his real arm and his paralyzed phantom arm.

Ramachandran describes the experiment in the following manner:

Philip rotated his body, shifting his shoulder, to “insert” his lifeless phantom into the box. Then he put his right hand on the other side of the

⁴³Bigna Lenggenhager, Tej Tadi, Thomas Metzinger, and Olaf Blanke, “Video Ergo Sum: Manipulating Bodily Self Consciousness,” *Science* 317, no. 5841 (2007).

⁴⁴V.S. Ramachandran and S. Blakeslee, *Phantoms in the Brain* (New York: William Morrow, 1998), 37ff.

mirror and attempted to make synchronous movements. As he gazed into the mirror, he gasped and cried out, “Oh, my God! Oh my God, doctor! This is unbelievable. It’s mind-boggling.” He was jumping up and down like a kid. “My left arm is plugged in again. It’s as if I’m in the past. All these memories from years ago are flooding back into my mind. I can move my arm again. I can feel my elbow moving, my wrist moving. It’s all moving again.

After he calmed down a little I said, “Okay Philip, now close your eyes.” “Oh, my,” he said clearly disappointed. “It’s frozen again. I feel my right hand moving, but there’s no movement in the phantom.” “Open your eyes” “Oh, yes. Now it’s moving again.”⁴⁵

Rubber hand illusions and phantom limbs both concern representation and modeling of body *parts*, but what about the body as a *whole* and the sense of the self related to the body as a whole? Could this also be manipulated experimentally? In other words, could one create a rubber-hand illusion of one’s entire body? Because the self is typically localized in the body, such an experiment would affect not only bodily attribution and location but self attribution and location as well—it would create the experience of being outside the physical boundaries of one’s actual body, an out-of-body experience. Metzinger and Blanke designed such an experiment with Bigna Lenggenheger and Tej Tedi. “When I first experienced the rubber-hand illusion, I immediately thought it would be important to see whether this would also work with a whole rubber body or an image of yourself. Could one create a full-body analog of the rubber-hand illusion? Could the entire self be transposed to a location outside the body?”⁴⁶ In the rubber hand illusion, bodily self representation was manipulated (specifically the sense of “mineness” or ownership of one’s body parts); the phenomenal

⁴⁵Ibid., 41.

⁴⁶Metzinger, *The Ego Tunnel : The Science of the Mind and the Myth of the Self*, 5.

self per se was not manipulated. In their “Video Ergo Sum” experiments, Lenggenhager, Tedi, Metzinger, and Blanke sought to manipulate attribution and localization of the entire body and to study the effects of such manipulations on one’s sense of self. In other words, the experiment sought to induce an out-of-body experience.

Participants were placed two meters in front of a video camera and fitted with a three-dimensional head mounted display, allowing them to view their backs in three-dimensional space. Metzinger described his own experience of wearing the head mounted display as viewing a 3-D version of Rene Magritte’s *La Reproduction Interdite*.⁴⁷ Participants’ backs were then stroked with a stick. Lenggenhager et al. varied the experimental conditions in several important ways. A time lag was introduced allowing for a discrepancy between seen and felt stroking. The virtual body was also varied. In the first trial, the participants viewed their virtual own body through the head-mounted display. In other trials, the camera angle focused on a mannequin and on a wood slab. Thus, participants viewed a virtual fake body and an object through the head-mounted display. Under synchronous conditions and with their virtual own body or virtual fake body, participants often felt as though the virtual body was their own body, actually identifying with it and “jumping into” it.⁴⁸ Metzinger describes his own experience in the following manner:

While I was looking at my own back as seen in the head-mounted display, Bigna Lenggenhager was stroking my back, while the camera was recording this action. As I watched my own back being stroked, I immediately had an awkward feeling: I felt subtly drawn toward my

⁴⁷Ibid., 98.

⁴⁸Lenggenhager, "Video Ergo Sum: Manipulating Bodily Self Consciousness."

virtual body in front of me, and I tried to “slip into” it. This is as far as things went.⁴⁹

Immediately after the stroking, participants were blindfolded, displaced and asked to return to their initial position. Participants showed a drift toward the virtual body in the synchronous condition with both the virtual own body and virtual fake body. The difference was weaker and no longer significant with the wooden slab. Lenggenhager et al. describe the significance of the experiment:

With the use of virtual reality and multi-sensory conflict, we induced an illusion that makes it possible to quantify selfhood by manipulating attribution and localization of the entire body. Our results show that humans systematically experience a virtual body as if it were their own when visually presented in the anterior extrapersonal space and stroked synchronously. This finding was corroborated by the participants’ mislocalization of their own bodies to a position outside their bodies, showing that self-attribution and localization of the entire body rely, at least partly, on similar visual-somatosensory integrative mechanisms to those of body parts.⁵⁰

While such experiments do not exactly duplicate OBEs, they are of tremendous significance because they show that the sense of self in relation to one’s own body can be manipulated under experimental conditions. Dualism is not a mysterious and intractable phenomenon but can be placed under the microscope, as it were, and what is being studied is not a soul but a phenomenal self-model. In the *Video Ergo Sum* experiments, persons or selves do not actually leave their own body and slip into the virtual body—all of this is the result of complex representational dynamics in the human brain. These experiments create something very similar to an OBE by creating multi-sensory conflict

⁴⁹Metzinger, *The Ego Tunnel : The Science of the Mind and the Myth of the Self*, 99-100.

⁵⁰Lenggenhager, "Video Ergo Sum: Manipulating Bodily Self Consciousness," 1098.

in a “top-down” manner. Such experiences can also be examined from the “bottom-up” as well.

As noted above, OBEs are part of a range of phenomena known as autoscopic illusions. Other types of autoscopic phenomena include: autoscopic hallucinations, feeling-of-a-presence, and heautoscopy.⁵¹ In autoscopic illusions, patients see their own body but they don’t identify with it and don’t have the feeling that they are “in” this illusory body. The feeling-of-presence is not a visual illusion but an illusion where the second illusory body is only felt. Heautoscopy is particularly fascinating. In these illusions the sense of self tends to alternate between two bodies or to be located between them. (The *Video Ergo Sum* experiments generated something very similar to a heautoscopy illusion).⁵² Blanke defines OBEs in terms of three fundamental characteristics: (1) Disembodiment: the feeling of being outside one’s physical body; (2) Parasomatic Perspective: the presence of an elevated visio-spatial perspective; (3) Autoscopy: The seeing of one’s own body from the elevated perspective.⁵³

Phantom limb phenomena, the rubber hand illusion, and the *Video Ergo Sum* experiments highlight the pragmatic, virtual character of self-modeling and bodily representation. Under conditions involving multisensory conflict people readily attribute fake body parts to themselves and identify with both a virtual own body and a virtual fake body, attributing and localizing the self outside the boundaries of their actual physical

⁵¹Blanke, "The Demystification of Autoscopic Phenomena: Experimental Propositions."

⁵²Lenggenhager, "Video Ergo Sum: Manipulating Bodily Self Consciousness."

⁵³Olaf Blanke and Shahar Arzy, "The Out-of-Body Experience: Disturbed Self-Processing at the Temporo- Parietal Junction," *Neuroscientist* 11, no. 16 (2005): 16.

bodies. Blanke uses the framework of multisensory conflict to study full-blown OBEs in a neurological context. His research into OBEs was launched when, seeking to locate the seizure site for an epileptic patient, he stimulated an OBE at the right angular gyrus.⁵⁴ An initial stimulation of 2.0-3.0 microamps resulted in the patient reporting that she was either sinking in the bed or falling from a height. When Blanke increased the stimulation above 3.0, the patient reported seeing herself lying in bed from above, seeing only her arms and legs. Blanke then had the patient focus on her arms and legs at 4.0-5.0 mA. She reported her legs and arms becoming shorter. The patient's limbs were bent at a 90 degree angle. When stimulated, the patient felt that her arms and legs were moving toward her face and tried to avoid them.

Recall the three defining characteristics of OBE described by Blanke: (1) disembodiment, (2) parasomatic perspective, and (3) autoscopia. Blanke argues that a two-fold disintegration is necessary to create an OBE. First, multisensory conflict and disintegration—the failure to bind proprioceptive, tactile, and visual sensations—is necessary to create the autoscopic aspect of the OBE—seeing one's own body outside its own physical boundaries. This sort of conflict was precisely what was induced in the rubber hand illusion experiments and the *Video Ergo Sum* experiments. Importantly, Blanke notes that a second form of conflict—vestibular conflict—is necessary to generate the sensations of disembodiment and parasomatic perspective. This conflict involves the relationship between external visual space and one's internal frame of reference created by vestibular information. “If the spatial frame of reference created by our sense of

⁵⁴Olaf Blanke, Theodor Landis, Laurent Spinelli, and Margitta Seeck, "Out-of-Body Experience and Autoscopia of Neurological Origin," *Brain* 127, no. 2 (2004).

balance and the one created by vision come apart, the result could be the conscious experience of seeing one's body in a position that does not coincide with its felt position."⁵⁵ Graviceptive and otholithic illusions resulting from vestibular disintegration have also been investigated scientifically. Such graviceptive-otholithic illusions frequently occur as responses to microgravity in space missions and on parabolic flights.⁵⁶ Significantly, altered perspective is one common manifestation. Inversion illusions, for example, involve 180 degree inversions of visual spatial perspective relative to extra personal space; persons experience themselves as if their bodies were inverted 180 degrees. Blanke argues that an analogous type of vestibular disintegration—along with multisensory disintegration—is necessary for OBE.

Research by Blanke has localized the neural correlate of OBEs at the temporal-parietal junction (TPJ). When healthy subjects were asked to imagine their bodies in the position of an OBE, the TPJ activated in less than half a second.⁵⁷ When that same region is inhibited by transcranial magnetic stimulation, healthy subjects cannot imagine themselves in the position of an OBE.⁵⁸ Importantly, other studies have highlighted the central role of the TPJ in multisensory processing, vestibular processing, and higher-level self processing: exactly the conditions Blanke points to as necessary for OBEs. TPJ along with cortical areas along the intraparietal sulcus have been associated with integrating

⁵⁵Blanke, "The Out-of-Body Experience: Disturbed Self-Processing at the Temporo-Parietal Junction," 20.

⁵⁶Ibid., 21.

⁵⁷Blanke, "Out-of-Body Experience and Autoscopy of Neurological Origin."

⁵⁸Ibid.

tactile, proprioceptive, and visual information. Further, core regions of the human vestibular cortex are situated at the TPJ, including the posterior insula, where brain damage is associated with graviceptive illusions. TPJ is also involved with perception and representation of the body. Finally, TPJ is also associated with higher-level self processing including egocentric visio-spatial perspective, agency, and the self-other distinction. “These data suggest an interaction between lower-level vestibular and multisensory processing and higher-level self processing such as egocentric visio-spatial perspective taking, agency, and self-location.”⁵⁹

Empirical research on OBEs provides support for the virtual character of the phenomenal self model and sheds a great deal of explanatory light on the origins of dualism. Such research casts significant doubt on the claim that OBEs provide direct empirical confirmation of dualism. On the contrary, phenomena like OBEs make perfect sense in light of the understanding of the conscious self developed in the previous section.

This section provided a naturalistic sketch of the self and showed how the conscious self can be explained in a bio-psychological vertically integrated manner. In doing so, it has also shed significant light on how the brain can create the sense of being outside of the brain (ordinary out-of-brain experience) and even outside of one’s physical body (extraordinary out-of-body experience). The virtual experiences of being outside one’s brain and outside one’s physical body are two powerful sources of dualist

⁵⁹Blanke, "The Out-of-Body Experience: Disturbed Self-Processing at the Temporo-Parietal Junction," 21.

intuitions. Higher-level cognitive science has also shed a great deal of light on dualism.

It is to this research that we now turn.

Toward a Bio-Psychological Explanation of Dualism:

Insights from Cognitive Science

Chapter One and the preceding section were concerned to establish consciousness and the conscious self as real but entirely natural phenomena, capable of being explained at the levels of scientific psychology and neuroscience. However, it is quite easy to think that the phenomenal self is something “over-and-above” the brain and the rest of the physical body. The transparency of the conscious self—the fact that all of our experience appears to be “out-of-brain” experience—helps to explain the intuition that selves are distinct from bodies. Folk phenomenology is one powerful source of dualist intuitions, but it is not the only one. In the course of their evolutionary history, human beings have not only developed an elaborate common-sense sense of self; they have also developed an elaborate common-sense knowledge of the other objects and “subjects” encountered in the world. Understanding this common-sense knowledge also contributes much to the explanation of dualism and other aspects of human religiosity. This section explains how.

Human beings possess complex and often implicit knowledge of material objects (folk physics), living things (folk biology), and other human minds (folk psychology). Cognitive-developmental psychology has uncovered a massive amount of information about these various inference systems, and in recent years, these findings have been applied fruitfully to the study of religion, leading to the emergence of the discipline known as the cognitive science of religion.

This section examines naturalistic explanations of dualism and theism from the vantage point of cognitive-developmental psychology. It begins with an overview of the findings of cognitive science on folk physics, folk biology, and folk psychology. It then applies these findings to questions concerning religious dualism.

Introducing Cognitive Science

In the previous section, the complex conscious autobiographical self that modern human beings possess was rooted in an evolutionary history involving the interaction between organism with an “internal milieu” and constantly changing external environment. That environment is filled with physical objects, living things, other animals, and most importantly, other human beings. Just as the complex sense of self possessed by modern human beings can be rooted in the fairly rudimentary notion of an “internal milieu,” so too, many forms of human knowledge about physical, biological, and social worlds can be understood to emerge primarily as adaptive skills acquired by organisms in particular evolutionary contexts.

To understand how this is the case, it is important to draw a crucial distinction between implicit and explicit knowledge. Knowledge is typically thought of in terms of explicitly held beliefs acquired through learning and education and conscious reasoning and deliberation. Many, in fact, have held the position that all knowledge is of this sort; indeed this is the foundation of the “blank slate” understanding of human nature discussed above. The cognitive revolution in psychology changed all of this by proposing that knowledge and beliefs can be implicit as well as explicit. Questions concerning the acquisition of language led to the demise of a behaviorist psychology focused on explicit

learning and ushered in the cognitive revolution in psychology. Noam Chomsky's groundbreaking work in linguistics in the 1950's called the most fundamental assumptions of behaviorism into question. Chomsky called attention to underlying similarities in the grammars of languages and to children's ability to learn language rapidly despite being exposed to it in a very piecemeal manner ("the poverty of the stimulus"). Chomsky argued that human beings possessed an inborn "language acquisition device" that guides very young children toward an effortless mastery of their native language and its grammatical rules.

Chomsky's pioneering research led to an explosion of work in the cognitive sciences. Fundamental to much of this work is the notion that human beings possess innate and implicit forms of knowledge that interact in various ways with explicit learning channels. We can effortlessly construct grammatically correct sentences in our native language without knowing any explicit rules of grammar. Explicit knowledge comes later with former education. Cognitive scientists have extended this fundamental insight well beyond the specific issue of language. As Justin Barrett puts it,

[c]ognitive scientists...have concluded that the adult human mind has a large number of devices that are used for different problems and on different occasions...Most of these mental tools operate automatically, without any conscious awareness. They efficiently and rapidly solve lots of problems without concentration or angst, much in the same way that computer programs solve problems in a swift, effortless fashion.⁶⁰

This insight is often expressed in terms of the notion of "domain-specificity": the notion that the human mind-brain employs different specialized inference systems in

⁶⁰Justin L. Barrett, *Why Would Anyone Believe in God?* (Walnut Creek, CA: AltaMira Press, 2004), 3.

response to objects in different conceptual categories. Contemporary cognitive science draws heavily on evolutionary theory and understands such inference systems to be in place as responses to problems recurrent in human environments for hundreds of thousands of years. Pascal Boyer provides a convenient summary of some of these essential principles:

- Perception and understanding of surroundings require inferences and guesses about different aspects of objects around us.
- The mind is composed of specialized systems that produce inferences about these different aspects.
- Objects in different “ontological categories” activate different sets of these specialized systems.
- Each inference system is itself composed of even more specialized neural structures.
- Inference systems make us attend to particular cues in environments and produce specific inferences from these cues.
- Skeletal versions of the principles direct knowledge acquisition from infancy.
- All concepts develop as skills, which is why discussions of innateness are often meaningless
- What principles you have depends on what species you are: which is why evolution is relevant to mental architecture.⁶¹

Inference systems leave no trace in the fossil record; evidence for their existence comes primarily from cognitive-developmental psychology: especially the study of very young children. Such studies are significant because they provide the best control for linguistic and cultural influences and thus provide an empirical means of adjudicating what were once entirely philosophical claims about the human person. As Edward Slingerland observes:

⁶¹Pascal Boyer, *Religion Explained : The Evolutionary Origins of Religious Thought* (New York: Basic Books, 2001), 106,126.

If assumptions or abilities can be clearly demonstrated in infants or prelinguistic children, we can be fairly confident that we have eliminated the influence of language and – at least in the case of infants – culture. Obviously, the linguistic and cultural assumptions of the experimenters are still relevant, but the fact that... child development results... have been replicated across a wide spectrum of cultures increases our confidence that we are observing species wide cognitive defaults.⁶²

One cannot talk to babies and babies cannot perform complex behavioral tasks, yet there is a way to effectively study their mental life. As developmental psychologist Paul Bloom notes:

Babies may have little control over their bodies, but they can willingly move their heads and eyes. And what a baby looks at can tell you something about how it sees the world. This is because babies are like adults in some regards. If they see the same thing over and over again, they get bored and look away. If they see something new or unexpected, they look longer. Thus, analyzing looking time can tell us what babies think of as being ‘the same thing’ and what they see as being ‘new or unexpected.’⁶³

Following such a methodology, studies in cognitive-developmental psychology have documented three especially important inference systems. These include common sense knowledge of material objects or folk physics, common sense knowledge of the living world or folk biology, and common sense knowledge of other minds or “folk psychology.”

In terms of common sense knowledge of physical objects or folk physics, extensive studies have documented that key assumptions about the behavior of such objects are in place long before children have extensive interaction with the material

⁶²Slingerland, 120-121.

⁶³Paul Bloom, *Descartes' Baby : How the Science of Child Development Explains What Makes Us Human* (New York: Basic Books, 2004), 8-9.

world. Elizabeth Spelke observes that both adults and infants assume that physical objects will operate according to four fundamental principles:

1. *Cohesion*: Objects are connected masses of stuff that move as a whole. If you want to know where the boundaries of an object are, an easy test is to grab some portion of stuff and *pull*—what comes with what you are pulling belongs to the same object; what remains does not.
2. *Solidity*: Objects are not easily permeable by other objects; if you tap at an object with your finger, your finger does not penetrate.
3. *Continuity*: Objects move in continuous paths; they travel through space without gaps. An object would violate this rule if it disappeared from one location and reappeared in another.
4. *Contact*: Objects move through contact. A ball on a pool table is not going to move unless something contacts it; it will not run from the cue or come when it is called. The exceptions to this rule are animate creatures, like people and dogs, and certain complex artifacts like robots and cars.⁶⁴

Human beings distinguish living and non-living things from an early age and attribute a unique set of properties to living things. Studies by Frank Keil, Scott Atran and others have documented that common sense knowledge of animals is driven by a kind of essentialism; animals are understood to have some internal property that is characteristic of the species and that cannot be removed.⁶⁵ Essences are linked to “insides” rather than to external appearances. Such essentialism is also closely linked with teleology, the notion that living things “possess properties that have purposes.” In terms of the understanding of religion, by far the most important form of common sense knowledge is folk psychology, especially agency detection and theory of mind.

⁶⁴Ibid., 6. See E. Spelke, "Initial Knowledge: Six Suggestions," *Cognition* 50 (1994).

⁶⁵See Frank Keil, *Concepts, Kinds, and Cognitive Development* (Cambridge, MA: MIT Press, 1989).

Folk Psychology: Agency Detection and Theory of Mind

Agency Detection

Common sense knowledge of the basic principles governing physical objects is clearly important for survival. Even more important, however, is knowledge of agents, especially other human beings. It is not difficult to see why this is the case from an evolutionary perspective. Objects, in the sense used here, are inert physical things like rocks, sticks, plants, and artifacts. If they respond to the world at all, it is in a purely mechanical fashion.⁶⁶ Agents, on the other hand, are beings like animals and other human beings capable of independently initiating action. Because of this agents call for special attention. From an evolutionary perspective and in terms of our implicit knowledge, agents are the things that matter most in an environment. Human beings must rapidly and accurately distinguish between objects and agents in the environment. Agency detection—the ability to distinguish between agents and objects in an environment—is thus a crucial aspect of folk psychology.

As noted above, agency detection depends on both speed and accuracy, but, in terms of accuracy, confusing an agent for an object (e.g., a snake for a coiled rope) and confusing an object for an agent (e.g., a coiled rope for a snake) do not involve equal payoffs. As long as one can rapidly readjust, the cost of “false positives” such as confusing or interpreting object as an agent is minimal.⁶⁷ As anthropologist Stewart Guthrie has pointed out, agency detection is an implicit form of Pascal’s Wager: from an

⁶⁶Todd Tremplin, *Minds and Gods: The Cognitive Foundations of Religion* (New York: Oxford University Press, 2006), 70.

⁶⁷Boyer, 145.

evolutionary perspective, “betting” on agents over objects is a wise choice.⁶⁸ False alarms with agency are evolutionarily advantageous. This becomes the default interpretative strategy of the agency detection device. “When in doubt about whether something is an agent, assume that it is. It’s better to have a fast device that occasionally gets it wrong than a slow device that is always accurate.”⁶⁹

One thing crucial to keep in mind in terms of agency detection is that traits that evolved for one purpose can be used for another purpose. As Paul Bloom notes, “a mind that has evolved to respond with sexual arousal in situations with actual people (adaptive) can respond the same way to pornographic movies (nonadaptive); a preference for sweet fruit (adaptive) can drive one to gorge on candy (nonadaptive).”⁷⁰ It is important to note that this distinction between adaptive and non-adaptive purposes applies to the cognitive realm as well. Cognitive anthropologists Dan Sperber and Scott Atran distinguish between what they call a cognitive function’s “proper domain” and “actual domain.” The proper domain of an agency detection device (ADD) involves “stimuli that track behaviors of agents including human beings... identifying animate beings as agents with goals and internal motivations, would allow our ancestors to anticipate goal directed actions of predators, prey, friends, and foe and profit from this in ways that enhanced hominid survival and reproductive success.”⁷¹ The actual domain of ADD involves any

⁶⁸See Stewart Guthrie, *Faces in the Clouds: A New Theory of Religion* (Oxford: Oxford University Press, 1993).

⁶⁹Tremlin, 70.

⁷⁰Bloom, 33.

⁷¹Atran, 60.

information in the organism's environment that satisfies the module's input conditions—whether or not the information is functionally relevant to ancestral task demands.⁷²

The “promiscuous” nature of human agency detection has been demonstrated in a number of psychological experiments. In the 1940s, F. Heider and S. Simmel presented a movie in which geometrical figures move in certain ways. People instinctively described the figures as if they were specific people with goals and desires.⁷³ Atran points to a number of experiments where mental states were inferred from “poor and fragmentary triggering experiences” including: interruptible movement toward a goal,⁷⁴ self-propulsion,⁷⁵ coordinated motion between subjects,⁷⁶ pointing,⁷⁷ eye gaze,⁷⁸ facial expression,⁷⁹ and interactive gesture and signaling.⁸⁰

⁷²Ibid.

⁷³F. Heider and S. Simmel, "An Experimental Study of Apparent Behavior," *American Journal of Psychology* 57, no. 2 (1944).

⁷⁴G. Csibra, G. Gergely, S. Biro, O. Koos, and M. Brockbank, "Goal Attribution without Agency Cues: The Perception of "Pure Reason" In Infancy," *Cognition* 72, no. 3 (1999).

⁷⁵D. Premack, "The Infant's Theory of Self-Propelled Objects," *Cognition* 36 (1990).

⁷⁶D. and A. Premack, "Origins of Social Competence," in *The Cognitive Neurosciences*, ed. Michael Gazzaniga (Cambridge, MA: MIT Press, 1995).

⁷⁷A. Leslie, "The Theory of Mind Impairment in Autism," in *Natural Theories of Mind*, ed. A. Whiten (Oxford: Blackwell, 1991).

⁷⁸Simon Baron-Cohen, *Mindblindness: An Essay on Autism and Theory of Mind* (Cambridge, MA: MIT Press, 1995).

⁷⁹M. Tomasello, R. Strosberg, and N. Akhtar, "Eighteen-Month-Old Children Learn Words in Non-Ostensive Contexts," *Journal of Child Language* 23 (1996).

⁸⁰S. Johnson, V. Slaughter, and S. Carey, "Whose Gaze Will Infants Follow?," *Developmental Science* 1(1998).

Theory of Mind

Agency detection merely distinguishes between agents and objects in the environment. The actual attribution of complex mental states like propositional attitudes such as belief and desire require a much more elaborate system of social cognition, one that is—at least in most respects—unique to human beings. These more complex inferences concerning other minds are referred to as mentalizing, mind reading, and theory of mind.⁸¹ While there are many social species, such elaborate mind reading is uniquely human and indeed at the very foundation of human culture. As Alvin Goldman observes:

Mentalizing may be the root of our elaborate social nature. Would there be language and discourse without mentalizing? Would the exquisitely coordinated enterprises of cultural life, the structures of love, politics, and games, be what they are without participants attending to the mental states of others? Would there be a human sense of morality without an understanding of what others experience, of what their lives are or might be like? The notion that mentalizing anchors the fabric of social life partly accounts for the profusion of interest in the subject⁸²

The development of social cognition and specifically theory of mind follows a developmental trajectory that has been the subject of intensive research for the past thirty years.⁸³ Such findings are intimately connected with findings concerning agency detection discussed in the previous section. Very young babies have important expectations concerning faces and hands. Andrew Meltzoff famously demonstrated that

⁸¹The scientific study of mentalizing was launched largely as a result of David Premack and Guy Woodruff's important 1978 paper, "Does the Chimpanzee Have a Theory of Mind."

⁸²Alvin I. Goldman, *Simulating Minds : The Philosophy, Psychology, and Neuroscience of Mindreading, Philosophy of Mind* (Oxford ; New York: Oxford University Press, 2006).

⁸³Examples of this developmental trajectory are from Bloom (2004).

infants can already respond to the cues of an experimenter a mere forty minutes from birth.⁸⁴ If the experimenter sticks out his or her tongue, the infant will respond in a similar fashion, an impressive fact given the fact that infants have never seen their tongues in a mirror but already know that this is the part of their body that corresponds to that of the experimenter.

Young babies have expectations associated not only with faces but also with hands. In one experiment, psychologists presented babies with a display containing two different objects; a hand then reaches for one of the objects. The location of the objects is then switched. Babies expect that the hand should switch locations, too; however, babies do not have this expectation for sticks.⁸⁵ Significant inferences concerning self and other minds begin to surface between the first and second years of life.

The gold standard for theory of mind is generally considered to be the ability to attribute false belief.⁸⁶ The attribution of false belief is tested in a number of experimental tasks. Perhaps the most well known task is the so-called “Sally-Ann Task.” The scenario for the test is as follows: Sally and Ann are together in a room that contains a basket and a box. Sally places her marble inside a basket and leaves the room. While she is gone, naughty Ann moves her marble to the box. Sally returns. Where will Sally look for the marble? Three-year-olds fail at this task, saying that she will look in the box.

⁸⁴A.N. Meltzoff and M.K. Moore, "Imitations of Facial and Manual Gestures by Human Neonates," *Science* 198 (1977).

⁸⁵Bloom, 15; A.L. Woodward, "Infants Selectively Encode the Goal Object of an Actors Reach," *Cognition* 69 (1998).

⁸⁶This suggestion concerning the significance of false belief in testing theory of mind was made independently by the philosophers Daniel Dennett, Jonathan Bennett, and Gilbert Harman in response to Premack and Woodruff's article, "Does a Chimpanzee Have a Theory of Mind?"

Four- and five-year-olds pass the test with much greater frequency, reflecting an ability to distinguish between mental representations and reality. This developmental pattern continues. The Sally-Ann task tests what has been called first-order theory of mind (involving second-order reasoning): “I *think* she *thinks* X.” What has been called second-order theory of mind (involving third-order reasoning): “I *think* she *thinks* that I/he/she *think(s)* X,” emerges significantly later, around age seven.⁸⁷

The question of how such mentalizing occurs has been the topic of intense debate in contemporary cognitive-developmental psychology and philosophy of psychology. Theory of mind is frequently used as a synonym for mind reading or mentalizing, but theory of mind is itself actually a scientific *theory* about how mindreading occurs. Advocates of this approach argue that mindreading is accomplished primarily by *theorizing*; hence, it is referred to as the *theory* theory of mindreading. Others argue that mindreading is primarily accomplished by putting oneself in another’s shoes, thereby simulating the mental states of others. This approach is known as the *simulation* theory of mindreading.

Some of the best evidence for theory of mind and insight into its operation comes from studying its absence. It is possible for theory of mind to be highly impaired or not to exist at all. A person can have an intact common sense understanding of the world of objects with little or no understanding of the social world. Many psychologists and neuroscientists understand autism as a form of “mindblindness.”⁸⁸ Human beings

⁸⁷See Jesse M. Bering, and Becky D. Parker, "Children's Attributions of Intentions to an Invisible Agent," *Developmental Psychology* 42, no. 2 (2006).

⁸⁸Most importantly Baron-Cohen.

typically have the opposite problem of autistics. Human social intelligence overshoots in a number of ways, and this is precisely where theory of mind research becomes relevant to religion.

Folk Knowledge and Religious Belief:

Introducing Cognitive Science of Religion

How does religion fit into the explanatory framework of cognitive-developmental psychology? Cognitive scientists typically avoid the claim that religious beliefs are directly adaptive. They typically understand religion as a by-product of other directly adaptive inference systems: folk physics, folk biology, and folk psychology. But the question of *how* exactly religion is a by-product is the subject of significant debate in the cognitive science of religion.

One particular sort of byproduct explanation draws on a strand of cognitive anthropology known as cultural epidemiology. As the name implies, cultural epidemiologists are concerned with isolating factors that make certain cultural concepts contagious and easy to transmit. According to this sort of explanation, religious concepts are best understood as *minimally counterintuitive* concepts. Such concepts are primarily cultural creations that exploit key aspects of our cognitive architecture in two important ways. First, human beings have certain intuitive expectations concerning the concepts OBJECT, ANIMAL, and PERSON based on their folk physical, folk biological, and folk psychological inference systems. Minimally counterintuitive concepts become minimally counterintuitive by violating exactly one ontological expectation (e.g., Soul/Ghost/Spirit

= PERSON + Counterintuitive Physical Properties).⁸⁹ This makes such concepts memorable. Research on memory has documented that such minimally counterintuitive concepts are more readily recalled than maximally counterintuitive concepts or intuitive concepts.⁹⁰ Secondly, while minimally counterintuitive, such concepts can still run on the inference system associated with the concept. Thus in this account, we will reason about souls/spirits/ghosts in the same way we reason about ordinary persons, barring the counterintuitive properties.

In this version of by-product explanation, religious beliefs are seen primarily as explicit cultural creations that draw upon the implicit expectations associated with our folk physical, folk biological, and folk psychological inference systems. There is much explanatory potential in the notion of minimally counterintuitive concepts, but such an approach leaves important questions unanswered. This is particularly true in terms of the question of dualism. The framework of “minimally counterintuitive concepts” helps explain the *transmission* of religious concepts, but can it really explain how such concepts *originated* in the first place? Why do dualist religious concepts—e.g., God concepts, afterlife beliefs—show up cross culturally?

In recent years another trajectory has emerged in the cognitive science of religion that understands certain aspects of religious beliefs as the direct by-product of our inference systems, especially folk psychology. In the minimally counter-intuitive

⁸⁹For discussions of minimally counterintuitive concepts see Atran, *In God's We Trust: The Evolutionary Landscape of Religion*; Barrett, *Why Would Anyone Believe in God*; Boyer, *Religion Explained*.

⁹⁰A. Norenzayan and S. Atran, "Cognitive and Emotional Processes in the Cultural Transmission of Natural and Nonnatural Beliefs," in *The Psychological Foundations of Culture*, ed. M. Schaller and C. Crandall (Hillsdale, NJ: Erlbaum, 2002).

understanding, religious beliefs are explicit beliefs that run on implicit inference systems such as folk psychology. In this other trajectory, certain foundational aspects of religious beliefs are themselves seen as forms of implicit belief/knowledge, that then interact with explicit beliefs and knowledge. This line of research has especially focused on the tendency to view both the biological and non-biological natural world as designed for a purpose and on the tendency for human beings to think in a dualist fashion and believe in the continuity of certain psychological functions after death.

Natural Sources of Religious Dualism III:

Paul Bloom and Jesse Bering on Folk Psychology and Dualism

Until recently, the topic of dualism per se has not figured prominently in discussions in the cognitive science of religion. But dualism is central to religion in at least two ways: (1) most religions hold dualist or idealist beliefs about human nature and its destiny after death, (2) other religious concepts are also premised on a mind-body dualism, i.e., God, gods, and spirits are also typically understood as “persons without bodies.” Dualism must be at least plausible in order to get religious beliefs off the ground. Dualism is essential not only for explaining soul beliefs but for explaining God beliefs as well.

Recent research by the developmental psychologists Paul Bloom and Jesse Bering has drawn attention to the role that our cognitive systems play in the production of common-sense dualism (a “folk psychology of souls”) and the role of that dualism in the production of religious belief. Their research suggests that dualism of a fairly robust sort

may in fact be the cognitive default of *homo sapiens*. This implicit dualism then interacts in various ways with different learning channels (e.g., science, religion).

Thus far we have seen that human beings have a number of distinct cognitive systems for navigating physical, biological, and psychological worlds. That these systems are separate and distinct was illustrated by conditions like autism where folk physics exists independently of a robust folk psychology. Paul Bloom argues that the fact that we have two separate *cognitive systems* for understanding objects and understanding minds leads us to see minds/souls and bodies as also separate and distinct.⁹¹ As Bloom notes, both of these systems are biological adaptations in the human brain working in a cold-bloodedly rational way to help people anticipate and understand physical and social reality. They are like two different computer programs running within the brain performing different tasks. Because we have two distinct cognitive systems and because they have incommensurable outputs, dualism emerges as an evolutionary by-product or accident.⁹² We think of human beings as having mind/souls and having bodies. As in the case of folk phenomenology, ideas about folk physics and folk psychology easily lead to folk ontology: souls are seen as really distinct from bodies. Folk physics and folk psychology are cognitive systems located in the human brain, but we don't think of them that way, we think in terms of bodies and souls.

The previous section showed that that the ability to differentiate agents and objects is in place quite early developmentally. There is some evidence that very young

⁹¹Paul Bloom, "Is God an Accident?," *The Atlantic Monthly* (2005); Paul Bloom, "Religion Is Natural," *Developmental Science* 10, no. 1 (2007).

⁹²Bloom, "Is God an Accident?"

infants do not recognize human beings as physical objects at all, and that recognizing human beings as partly physical beings may in fact be a developmental accomplishment. They may be “superdualists” or idealists. In a study of five-month-olds, Bloom and his colleagues Valerie Kuhlmeier and Karen Wynn investigated whether infants apply folk physical principles like cohesion, solidity, and continuity to human beings: “do infants recognize that, in some cases, the principles they readily apply to inanimate objects also apply to humans?”⁹³ Kuhlmeier, Bloom, and Wynn focused specifically on the principle of continuity, drawing on research by Elizabeth Spelke and her colleagues demonstrating that infants expect continuity with inanimate objects. Infants express surprise when an object disappears behind a barrier and then seems to reemerge from behind a non-adjacent barrier. Kuhlmeier, Bloom, and Wynn performed an experiment that tested continuity expectations regarding persons rather than objects. Significantly, they discovered that the principle of continuity was not applied to persons; infants were not surprised when a human being seemed to violate the law of continuous motion. Reflecting on these different expectations concerning objects and persons, Kuhlmeier, Bloom, and Wynn write:

It is possible that the dissociation between objects and humans found in the present study forms the complement to the distinction between human and inanimate objects in terms of social, goal-directed behavior. Together, this double dissociation suggests that young infants may have different modes of construal for humans versus inanimate objects: humans are construed in terms of social and intentional actions, while inanimate objects are interpreted via a system sensitive to object physics....The existence of this human/inanimate distinction, and the differential

⁹³Valerie A. Kuhlmeier, Paul Bloom, and Karen Wynn, "Do 5-Month-Old Infants See Humans as Material Objects?," *Cognition* (2004): 95.

application of principles to each, may help infants to define these areas of knowledge early in development. The appreciation that these construals overlap—that in certain regards, people *are* just objects—may be a developmental accomplishment.⁹⁴

Independent work by developmental psychologists Jesse Bering and Paul Harris has opened up interesting lines of research pertaining specifically to dualism and afterlife beliefs. The experiments by Bering, Harris and their colleagues all involved scenarios involving the death of an agent. Participants were then asked a series of questions regarding the continuity of various mental states and biological functions after death. As Harris notes, such studies have direct bearing on the question of dualism:

If children think of the person as an indivisible whole, then whether we ask them about bodily processes (e.g. the functions of the eyes or the brain) or about concomitant mental processes (e.g. the functioning of sight or the mind) they should give consistent replies—claiming either that bodily and mental processes have stopped or that both types of process continue. On the other hand, if children are susceptible to dualist thinking—to construe mental processes as separate from, and independent of bodily processes—then they would likely offer a different pattern of replies in the two cases, claiming, for example, that bodily processes no longer function whereas mental processes continue.⁹⁵

Work by Bering and his collaborators and independent work by Paul Harris and his collaborators show that children are much more likely to understand psychological functioning continuing at death than biological functioning. This was even true in cases where a certain bio-psychological function (e.g., vision) was broken into biological (eyes) and psychological (seeing) components. Bering and Bjorklund found that children are

⁹⁴Ibid., 8.

⁹⁵Paul L. Harris, and Marta Gimenez, "Children's Acceptance of Conflicting Testimony: The Case of Death," *Journal of Cognition and Culture* 5, no. 1-2 (2005): 7-8

much more likely to see the mental process continue than the biological process.⁹⁶ Harris and Marta Gimenez discovered a similar differentiation of mind and body with Catholic children in Spain, and Harris and Rita Astuti found a similar pattern working with Vevo children in rural Madagascar.⁹⁷ Paul Harris describes the significance of these joint findings:

How far are our conclusions and interpretations consistent with other research on afterlife beliefs and their development? A robust finding, both in Study 1 and in the earlier report of Harris and Gimenez (2005), was that participants are more likely to acknowledge the continued functioning of mental as compared to bodily processes after death. Bering and his colleagues...reported a similar differentiation between mind and body. Across several experiments, participants were more likely to assert the continued functioning of mental (e.g. feeling desires or emotions) as compared to biological (e.g. brain functioning) or psychobiological processes (e.g. feeling sick).⁹⁸

The Bering and Harris groups agree *that* both children and adults are dualists; however, there is substantial disagreement between the groups concerning the explanation of this dualism. At stake is the issue of how common sense psychological and biological knowledge relate to one another and to various learning channels (religious, scientific, etc.) during development. Harris and his colleagues are principally concerned with how children reason about death in secular and religious contexts and favor an explanation of afterlife beliefs that is primarily cultural. Their approach draws on Pascal Boyer's notion that religious beliefs typically involve minimally

⁹⁶Bering, "The Natural Emergence of Reasoning About the Afterlife as a Developmental Regularity."

⁹⁷Rita Astuti, and Paul Harris, "Understanding Mortality and Life of the Ancestors in Rural Madagascar," *Cognitive Science* 32 (2008).

⁹⁸Ibid., 736.

counterintuitive concepts. Astuti and Harris make the case that a consolidated understanding of biological death may in fact contribute to the acquisition of afterlife beliefs, precisely because of their counterintuitive nature.⁹⁹

This counter-intuitive concept explanation is likely part of the story, especially where *explicit* afterlife beliefs are concerned, but a number of questions can be raised about this explanatory strategy. First, this explanation states that children (and adults) acquire afterlife beliefs through culture. Children and adults are prone to acquire such concepts precisely because they violated biological expectations concerning death. The minimally-counter intuitive nature of afterlife concepts makes them especially “catchy.” This may be a plausible account of how explicit afterlife beliefs are *acquired*, but it leaves unexplained the question of why dualist afterlife beliefs show up within every human culture. It does not explain the etiology of afterlife concepts and beliefs *themselves*. How does the notion of a “person without a body” or an immaterial mind concept get there in the first place? As Bering notes:

[T]he question remains why individuals the world over represent the minds of those who have died to be active *at all*, since tacit biological knowledge of death and agency might just as easily have resulted in the complete absence of the concepts *ghost, spirit, or afterlife*... While cultural epidemiologists might presuppose that it is precisely this intuitive knowledge of death renders the child susceptible to afterlife concepts... this does not solve the very real problem of why specific counterintuitive concepts such as those dealing with the afterlife would spontaneously happen to appear in nearly all societies to begin with.¹⁰⁰

⁹⁹Ibid.

¹⁰⁰Jesse M. Bering, "Intuitive Conceptions of Dead Agents' Minds: The Natural Foundations of Afterlife Beliefs as Phenomenological Boundary," *Journal of Cognition and Culture* 2, no. 4 (2002): 268-269.

Research by Bering sheds some interesting empirical light on these questions, suggesting that implicit knowledge may be at work in the formation of afterlife beliefs. Bering's work trades on the distinction between explicit and implicit cognition introduced earlier. He is not claiming that explicit beliefs like "the soul goes to heaven when you die" are innate, but rather that our innate social cognition biases us toward certain kinds of afterlife belief which then interact with various learning channels.

[I]s it possible that the general idea of an afterlife is not so much implanted in people's heads by way of "exposure" to counterintuitive tales, as it is *already present*, already firmly entrenched in representational structures endemic to human cognition, and only then conceptually enriched through cultural channels? Can we not scarcely help *but* believe in some form of psychological continuity with the dead?¹⁰¹

Bering investigates this claim in several experiments. In "Intuitive Conceptions of Dead Agents' Minds: The Natural Foundations of Afterlife Beliefs as Phenomenological Boundary," Bering investigates whether people's reasoning about the minds of dead agents is isomorphic with their explicit beliefs about the afterlife and whether afterlife beliefs resemble the minimally counterintuitive template suggested by Pascal Boyer.

In *Religion Explained*, Boyer treats dead agents' minds as a specific form of minimally counterintuitive concept. For Boyer, a religious concept, as minimally counterintuitive, "preserves all the relevant default inferences except the ones that are explicitly barred by the counterintuitive element."¹⁰² He illustrates this with the concept of ghost or spirit:

¹⁰¹Ibid., 269.

¹⁰²Boyer, 73.

A good illustration is the familiar concept of ghost or spirit. This is found more or less the world over, not just in Gothic novels and Victorian séances. The concept is that of a PERSON who has counterintuitive physical properties. Unlike other persons, ghosts can go through solid objects like walls. But notice that apart from this ability, ghosts follow very strictly the ordinary intuitive concept of PERSON. Imagine a ghost suddenly materializes in your home as you are having dinner. Startled by this sudden appearance, you drop your spoon in your plate of soup. In a situation like that, your mind creates a whole lot of assumptions of which you are not necessarily conscious. For instance, you assume that the ghost *saw* you were having dinner, so she now *knows* you were eating. Also, the ghost probably *heard* the sound of your spoon landing in the soup and can now *remember* that you dropped it. You assume that the ghost *knows* you are here, since she can *see* you. It would be unsettling but not too surprising if the ghost asked you whether you were enjoying your dinner. It would be very weird if she asked you why you never had dinner at home or why you never had soup. In other words, you assume that this ghost has a mind. All of the italicized verbs above describe the sort of thing a mind does: it perceives actual events in the world and forms beliefs on the basis of those perceptions.¹⁰³

Do people actually represent the minds of dead agents in the manner suggested by Boyer? Do they represent minds of dead agents simply according to their explicit beliefs about the afterlife? Bering argues against these approaches. “The postulate that *ghost* is a cultural invention may be misguided in that it overlooks the possibility that not only the spontaneous inference system triggered by it, but also the very concept itself, has natural foundations in the human mind.”¹⁰⁴ This more nativist line of reasoning is not the absurd notion that people come ready equipped with the notion that the soul returns to heaven when they die. It is rather the notion that such explicit conceptions of the afterlife piggyback on more implicit conceptions of afterlife.

¹⁰³Ibid., 73-74.

¹⁰⁴Bering, 269.

Bering argues that these more implicit notions of afterlife are generated largely through the operation of theory of mind. This is Bering's *simulation constraint hypothesis*. As discussed earlier, there is debate among developmental psychologists on how mentalizing or mind reading takes place. One account holds that mind reading takes places largely through simulation, or putting oneself in the shoes of another person, imagining what it is like to be in their particular mental state. According to this account, passing false belief tasks involves the children's ability to put themselves in the shoes of another person whose knowledge is different from their own. Simulation thus depends on our own conscious experience of "knowing what it is like" to be in certain mental states.

Crucial to Bering's argument is the observation that death places impossible constraints on such simulation. We cannot imagine what it is "like" to be dead because this involves consciously representing a final state of unconsciousness. Further, simulation is not uniform across all mental states. There are some states that it is easier-to-imagine-the-absence-of (EIA states) than others, difficult-to-imagine-the-absence-of-states (DIA states). The absence of certain mental states appears readily in our conscious experience. We know what it is like not to be hungry or not to see or hear, because such mental states are readily switched on and off in our conscious experience. It is much more difficult to simulate the absence of thinking or emotion. Based on the fact that it is impossible to consciously simulate death (i.e., to consciously simulate permanent non-consciousness) and that some mental states are easier to imagine the absence of than others, Bering hypothesizes that there will be a tendency to project DIA states onto dead

agents' minds. Because we cannot simulate what it is like to be dead, we often attribute states that are difficult to imagine being without to dead agents.

Bering argues there are factors other than simulation constraints at work in the generation of implicit afterlife beliefs. The genesis of afterlife beliefs can also be understood through a *theory* theory approach to mindreading. The essential point is that whether one relies on simulation or not, reasoning about dead agents' minds is a difficult task because individuals simply cannot use the same explanatory frame used in the regular social domain. According to *theory* theory, one's theory of mind is developed by testing and refining hypotheses based on the behavior of others. Dead agents' minds present obstacles here as well, because their behavior provides limited information about underlying mental states. As Bering notes, the absence of action does not necessarily imply the absence of mental states. Young children, for example, tend to attribute mental states to agents in a dreamless sleep.¹⁰⁵

Along with these explanations based on simulation and theory approaches to mindreading, Bering suggests that another aspect of our social cognition may be responsible for the genesis of afterlife beliefs: the off-line character of much of our social reasoning.¹⁰⁶ Human beings engage in social reasoning not only when others are actually present, but in their absence as well. This leads us to assume that the individuals with whom we have relationships are engaged in actions even when we cannot directly observe them. This extends not only to living agents but to dead agents as well.

¹⁰⁵Bering, "The Folk Psychology of Souls," 455.

¹⁰⁶Ibid., 455-465.

Additionally, certain mental states involving external sensory organs seem to us more directly connected to biological functioning than, say, thinking or desiring. Such perceptual and psychobiological states seem much less ephemeral than thought and emotion, which seem to have no obvious connecting to physicality. Since such states seem more intimately connected with our biology and physicality, they are more prone to become integrated into our folk biological knowledge that death is the end of the organism. Mental states with no directly observable connection to our biology/physicality will be less prone to such integration.

Is this DIA/EIA distinction in fact evident in people's reasoning about the afterlife? As Bering notes:

In order for the simulationist model of the "intuitiveness" of belief in life after death to be borne out, several hypotheses must be supported. First, if indeed DIA states are more resistant to discontinuity reasoning because this requires that explicit biological knowledge must deliberately "turn off" the default simulation system, then all individuals, no matter their professed beliefs on the fate of personal consciousness after death, should be more likely to say that DIA states continue after death than they are EIA states, where biological knowledge can effectively co-opt the default system....In addition, it could be argued on the basis of the predictions made by the simulation model that discontinuity response latencies should be, in general, longer for DIA than for EIA states, since the former would require that reflexive simulation attempts be overwritten by explicit biological theorizing to arrive at such responses, reasoning about death in full scientific fashion, while the latter employs default simulations.¹⁰⁷

On the other hand, Bering notes that if a "culture only" explanation of afterlife beliefs is correct, no differentiation should show up between DIA/EIA in people with

¹⁰⁷Bering, "Intuitive Conceptions of Dead Agents' Minds: The Natural Foundations of Afterlife Beliefs as Phenomenological Boundary," 274.

differing afterlife beliefs.¹⁰⁸ People's understanding of the afterlife will reflect their explicitly learned beliefs. For example, an extincivist will say that all processes stop at death. A reincarnationist will say that all processes continue. An immortalist who understands the soul to leave the body at death, will understand it in a manner similar to Boyerian spirit concepts, as possessing the same mental properties it did before death.¹⁰⁹

Bering tested this hypothesis on undergraduate students. The students were presented with a story about an individual meeting an unexpected death. Participants were then asked questions about the individual in the story and instructed to answer the questions seriously. They were also told that there was no right or wrong answer; the experimenter was only interested in their opinion.¹¹⁰ All questions were prefaced with "Now that he/she is dead..." Interviews contained questions dealing with the following categories: BIOLOGICAL, PSYCHOBIOLOGICAL (Is he still *hungry*?), PERCEPTUAL (Can he *see* the paramedic trying to resuscitate him?), EPISTEMIC (Is he still *thinking* about his wife?), EMOTIONAL (Is he still *angry* with his wife?), DESIRE (Does he *want* to be alive?).¹¹¹

At the very end of the interview (so as not to bias responses), participants were asked to classify their explicit afterlife beliefs according to five categories:

Extincivist: Personal consciousness ceases entirely at death.

Agnostic: Uncertain as to what becomes of personal consciousness after death.

¹⁰⁸Ibid., 275.

¹⁰⁹Ibid.

¹¹⁰Ibid., 275-277.

¹¹¹Ibid., 276-277.

Immortalist: Personal consciousness survives the death of the body and exists forever.

Reincarnationist: Personal consciousness is reincarnated into a new body after death.

Eclectic: Combination of immortalist and reincarnation classifications.

Other: Personal consciousness survives the death of the body but uncertain what happens after this.¹¹²

The findings of the experiment support Bering's hypothesis that factors other than explicit knowledge are at work in the formation of afterlife beliefs. All groups of participants stated that biological and biopsychological functions ended with death; all groups also showed a significant drop in discontinuity responses for the emotional, desire, and epistemic categories. The major difference in response patterns occurs between agnostics, extinctivists, and all other groups. Agnostics and extinctivists had significantly higher discontinuity responses in the emotional, desire, and epistemic categories. Significantly, however, these discontinuity responses in those categories were not at ceiling. The four other groups produced very high discontinuity responses for biological and psychobiological categories; the rate of discontinuity responses dropped significantly for the perceptual category, and approached floor levels for the emotion, desire, and epistemic categories.

Such findings are significant because they suggest that dead agent's minds are represented in a manner different from the minimally counterintuitive concept model (Spirit = Invisible Person) suggested by Pascal Boyer. As Bering observes:

In terms of the cognitive underpinnings of religion, contrary to previous accounts (cf. Boyer, 2001), most afterlife believers do not represent ghosts or spirits simply as invisible human beings, but rather as invisible human beings with a narrower range of subjective experience than living agents –

¹¹²Ibid., 278-279.

experience that is delimited by the inactivation of *specific* psychological systems at death.¹¹³

The pattern of response offered by reincarnationists and extinctivists also supports the hypothesis that more is at work in reasoning about dead agent's minds than explicit beliefs about the afterlife.

[A]lthough they were the lowest of all, one would have expected the reincarnationists to have provided an even lower percentage of discontinuity responses for the biological and psychobiological questions than they did (83% and 90% respectively), only if to support their stated explicit beliefs that "personal consciousness" becomes reincarnated into another physical body at some point after death. Likewise, given their adamancy on the questionnaires that personal consciousness is exhaustively terminated at death, it is somewhat surprising that the extinctivists were not at ceiling with discontinuity responses for the DIA mental states (i.e., emotional, desire, and epistemic states; 68%, 68%, and 64% respectively), and in fact were significantly less likely to state that epistemic states discontinued after death than they were the two EIA states. Theoretically, if explicit (i.e., learned or adopted) beliefs were the whole story, reincarnationists should have been at absolute floor for all of the questions posed to them, and extinctivists at absolute ceiling.¹¹⁴

The simulation constraint hypothesis is also borne out in measurements of response latencies for the various questions types. Participants required more time to state that DIA states did not continue after death than they did for EIA states.

Bering and his colleagues have also have explored afterlife beliefs in children. If a purely cultural account is true, then the youngest children should exhibit the lowest number of continuity responses and continuity responses should increase with age. If other factors are involved, as the simulation constraint hypothesis suggests, belief in the

¹¹³Ibid., 288.

¹¹⁴Ibid., 289.

continuity of mental states should actually *decrease* with age. In experiments with American school children, Bering and David Bjorklund tested two hypotheses:

[I]f all that were influencing people's afterlife judgments were mainstream cultural beliefs, with increasing age, children should be more likely to make attributions of psychological continuation following death because of their increasing exposure to cultural norms.

Our first hypothesis was therefore that people's judgments about the continuity of psychological processes following death should actually *decrease* with increasing age, despite the background of cultural influence to the contrary....

Our second hypothesis was that these judgments about the continuity of psychological processes following death would also vary as a function of the nature of the mechanisms under question.¹¹⁵

Bering and Bjorklund presented students with a narrative in which a mouse is eaten by an alligator. The children were then asked a number of questions concerning the continuity of the various states introduced earlier: BIOLOGICAL, PSYCHOBIOLOGICAL, PERCEPTUAL, EPISTEMIC, EMOTIONAL, DESIRE. In experiments with children in the United States and with Catholic and secularly schooled children in Spain, Bering and Bjorklund found the same developmental trajectory: psychological continuity responses were, in fact, highest for the youngest children and decreased with age (Figure 8).¹¹⁶ This provides evidence that factors other than explicit religious indoctrination are responsible for afterlife beliefs. Bering argues that such a decrease is the result of increased biological knowledge and its relationship to psychological states.

¹¹⁵Bering, "The Natural Emergence of Reasoning About the Afterlife as a Developmental Regularity," 218.

¹¹⁶Ibid.

The findings show that older children and adults were, overall, more likely to state that both biological imperatives and psychological states ceased at death and were more likely to report that *particular* categories of psychological states (i.e. psychobiological and perceptual states) ended at death than that other did (i.e. emotional, desire, and epistemic states). In contrast the youngest children in the sample, although acknowledging that biological imperatives no longer applied to the dead agent, failed to distinguish between the different categories of psychological states and were just as likely to report that one type of state (e.g. psychobiological) continued after death as that another one did (e.g. epistemic). The finding may be explained by the implicit nature of the youngest children's knowledge about the biology of death; only after this knowledge has become conceptually enriched and made explicit can it be applied when reasoning about the psychological states of dead agents. However, even when explicit knowledge is in place, reasoning that certain types of psychological states (i.e. emotional, desire, and epistemic states) become extinguished at death appears difficult. Discontinuity reasoning for these types of mental states encounters resistance.¹¹⁷

As discussed above, Bering offers several explanations of this resistance:

simulation constrains, the off-line nature of social reasoning, and the ephemeral nature of these states. Interestingly, such a pattern was found in both religious and secularly schooled children in Spain. The Catholic students had higher levels of continuity responses, but followed the same developmental trajectory: the youngest students offered the greatest number of continuity responses. In terms of the higher overall continuity responses, Bering argues that explicit religious education might in fact reinforce the default cognitive stance in the face of explicit biological knowledge. It is in understanding the interaction between implicit afterlife beliefs and explicit afterlife beliefs acquired through religious indoctrination that the notion of minimally counterintuitive concepts becomes helpful. Bering hypothesizes that explicit afterlife

¹¹⁷Ibid., 228.

beliefs might become especially salient precisely because they violate biological expectations about death.

One of the roles served by religious instruction concerning 'life after death' may be to conceptually enrich a default stance that involves attributing mental states to dead agents. Although the present study did not directly test for this, religious pedagogy on the topic of the afterlife might serve to quarantine biologically relevant information such that this information is not used for theoretical applications concerning the fate of agents' minds after death. It is important to clarify that we are not arguing that biological knowledge among religiously schooled children is more impoverished than that of their secularly schooled peers, but only that these children may encounter a greater degree of conflict, and hence separation between biological verities and their religious beliefs. As such, biological reasoning about the psychological status of dead agents may be set aside in favour of explicit religious ideas that defy naturalistic principles and that are regularly communicated to these children by adult authority figures... Indeed, it may be precisely this disparity between biological facts and religious ideas that contributes to the memorability of the latter, fostering their cultural transmission to children... This was supported by the present data in that, at every age, there were fewer Catholic school students who could be classified as consistent cessation theorists than there were secular school students who had such a response profile.¹¹⁸

The cognitive science of dualism is significant not only in shedding light on the dualist view of human nature and the afterlife beliefs that many religions share. Such dualism is also important in understanding the origin of God concepts and God beliefs. God is understood very much on the model of an immaterial person. Without an initial sense that minds and bodies are distinct, belief in God would be fundamentally unintelligible. As Paul Bloom writes:

Most significantly for religion, dualism makes it possible to imagine souls without bodies. Christianity and Judaism, for instance, involve a God who

¹¹⁸Jesse M Bering, Carlos Hernandez Blasi, and David F. Bjorklund, "The Development of 'Afterlife' Beliefs in Religiously and Secularly Schooled Children," *British Journal of Developmental Psychology* 23, no. 4 (2005): 600.

created the universe, performs miracles, and listens to prayers. He is omnipotent and omniscient, possessing infinite kindness, justice, and mercy. But he does not, in any literal sense, have a body.¹¹⁹

But factors other than dualism are involved in the production of God concepts and beliefs, most importantly the over-attribution of agency and design and the tendency to view events as symbolic communications.

Natural Sources of Religious Dualism IV: From Dualism to Theism:

Over Attribution of Agency and Design

Over Attribution of Agency

It is not difficult to see the implications of agency detection for the study of religion. As Justin Barrett notes, agency detection contributes to the formation of religious concepts in a number of different ways. It can both contribute to the origin of religious beliefs and reinforce and encourage religious beliefs already held.

Sometimes HADD's [Hypersensitive Agency Detection Device] tendency to attach agency to objects contributes to the formation of religious concepts. The most straightforward manner is in identifying some ambiguous thing, such as a wispy form as an intentional agent—a ghost. With the assistance of face detectors and other tools sensitive to human forms, occasionally people see what appear to be humanlike figures ... Whether the sighting is an illusion or not, if the right information is fed to these mental tools, the outcome is a nonreflective belief in a ghost or spirit. Without sufficient reflective defenses, this nonreflective belief will become a reflective one.¹²⁰

However, agency need not be directly attributed to an object. A known object may be manipulated or controlled by an intentional agent acting “behind the scenes”:

¹¹⁹Bloom, "Religion is Natural," 149.

¹²⁰Barrett, *Why Would Anyone Believe in God?*, 33.

HADD's tendency to find agency in objects contributes to the formation of religious concepts in a second manner. Often the objects that HADD registers as being agents are known objects. Unlike in the case of spirits, HADD may suggest that known nonagents are exhibiting agency. A storm cloud might have destroyed one and only one home in a village with hail and lightening. Under some conditions, HADD might register the cloud as an agent acting purposefully. But a cloud is not an agent... [T]hough HADD may have detected an object behaving like an agent, a more salient candidate may be attributed responsibility for the action in question. For instance, if villagers believe a certain god controls the weather, the storm cloud's apparent agency might be directed by that god against the reprobate individual. In these cases, HADD encourages belief in already known superhuman agents.¹²¹

Over Attribution of Design

We are examining the origins of immaterial mind concepts like God and the soul and belief in these concepts. One major source of intuitions about the existence of God—i.e., a divine mind—is design, the sense that the universe as a whole and the organisms in it were designed for a purpose. Design arguments are the most commonly cited justification for belief in God.¹²² Proposals to teach the so-called “Intelligent Design Theory” as part of the biological curriculum in public high schools generated a national debate over a century after Darwin. Anthropic “fine tuning” arguments remain popular among some very sophisticated thinkers. Important work by the cognitive-developmental psychologist Deborah Kelemen indicates that the tendency toward “promiscuous teleology” may in fact be the cognitive default, a by-product of over-zealous social cognition. Theory of mind is naturally extended not only to human beings but to the artifacts that human beings create. Research by Kelemen indicates that theory of mind

¹²¹Ibid., 33-34.

¹²²See Michael Shermer, *How We Believe: Science, Skepticism, and the Search for God* (New York: Henry Holt, 2000).

extends well beyond the world of human beings and their artifacts; it extends to the biological and non-biological natural world as well. Kelemen's research has documented children's tendency to understand objects as existing for a purpose, a bias which she dubs "promiscuous teleology."¹²³ Kelemen's work along with the independent research of M. Evans has also documented children's tendency to view natural phenomena as intentionally created by a non-human agent.¹²⁴

Research on teleological reasoning originally occurred in the context of discussions of common sense or folk biology. In that context, "studies have discovered that children's reasoning about living things is constrained by teleological assumptions from a very early age."¹²⁵ Kelemen's work is significant in finding that children apply such teleological reasoning to the non-biological natural world as well.

[W]hen asked to identify unanswerable questions, American 4- and 5-year-olds differ from adults by finding the question "what's this for?" appropriate not only to artifacts and body parts, but also to whole living things like lions ("to go in the zoo") and nonliving natural kinds like clouds ("for raining")....

These kinds of promiscuous teleological intuitions persist into elementary school, particularly in relation to object properties. For

¹²³See Deborah Kelemen, "The Scope of Teleological Thinking in Preschool Children," *Cognition* 70, no. 3 (1999); Deborah Kelemen, "Why Are Rocks Pointy? Children's Preference for Teleological Explanations of the Natural World," *Developmental Psychology* 35, no. 6 (1999); Deborah Kelemen, "British and American Children's Preferences for Teleo-Functional Explanations of the Natural World," *Cognition* 88, no. 2 (2003); Deborah Kelemen, "Are Children 'Intuitive Theists'?: Reasoning About Purpose and Design in Nature," *Psychological Science* 15, no. 5 (2004).

¹²⁴See E.M. Evans, "Beyond Scopes: Why Creationism Is Here to Stay," in *Imagining the Impossible: The Development of Magical, Scientific, and Religious Thinking in Contemporary Society*, ed. C.N. Johnson, K.S. Rosengren, and P.L. Harris (Cambridge, UK: Cambridge University Press, 2000); E.M. Evans, "The Emergence of Beliefs About the Origin of Species in School-Age Children," *Merrill Palmer Quarterly* 46 (2000); E.M. Evans, "Cognitive and Contextual Factors in the Emergence of Diverse Belief Systems: Creation Versus Evolution," *Cognitive Psychology* 42, no. 3 (2001).

¹²⁵Kelemen, "Are Children 'Intuitive Theists'?: Reasoning About Purpose and Design in Nature," 295.

instance, when asked to conduct a “science” task” and decide whether prehistoric rocks were pointy because of a physical process (e.g., “bits of stuff piled up for a long period of time”) or because they performed a function, American 7- and 8-year-olds, unlike adults, preferred teleological explanations whether they invoked “self-survival” functions (e.g. “so that animals wouldn’t sit on them and smash them”) or “artifact” functions (e.g. so that animals could scratch on them when they got itchy”) (Kelemen, 1999c).¹²⁶

Where do such teleological explanations come from? A culture-only explanation suggests itself. American children are teleologists because American culture is saturated with theism and theism of a very creationist sort. Kelemen argues against this explanation. The bias toward teleological reasoning in American children has been replicated in the much more secular context of Great Britain.¹²⁷ More significantly, independent work by M. Evans found that children from both fundamentalist and non-fundamentalist homes show a bias toward “creationist” accounts of how species originate.¹²⁸ Kelemen argues that such teleological and creationist reasoning may be the by-product of our over-active social cognition: “Perhaps children’s generalized attributions of purpose are, essentially, side effects of a socially intelligent mind and is, therefore, oriented toward explanations characterizing nature as an intentionally designed artifact—an orientation given further support by the artifact-saturated context of human cultures.”¹²⁹

¹²⁶Ibid., 295-296.

¹²⁷Kelemen, "British and American Children's Preferences for Teleo-Functional Explanations of the Natural World."

¹²⁸Evans, "Cognitive and Contextual Factors in the Emergence of Diverse Belief Systems: Creation Versus Evolution."

¹²⁹Kelemen, "Are Children "Intuitive Theists"?: Reasoning About Purpose and Design in Nature," 296.

This “intuitive theism” or “theism as by-product” thesis—that children make sense of the natural world in terms of a “somebody who designed them for a purpose” depends on several prerequisites being in place. (1) The capacity to maintain a mental representation of such an agent despite its intangibility, (2) the ability to attribute to that special agent mental states distinguishing it from more commonplace agents, (3) the ability to attribute design intentions to agents and understand an objects’ purpose as deriving from such intentions.¹³⁰ Kelemen argues that all three prerequisites are in place around age five, thus bolstering her case for intuitive theism.

In terms of the first prerequisite, children’s tendency to form imaginary relationships is indicative of their ability to reason about invisible individuals.¹³¹ Recent research has also highlighted children’s ability to distinguish between God and other agents when attributing mental states. Justin Barrett and his colleagues devised an experiment to test this, drawing on the shift between three- and five-year olds ability to pass false belief tasks.¹³² Barrett and his colleagues used a standard form of the test. Children were shown a cracker box and asked what it contained and then allowed to look inside the box, seeing that the cracker box actually contained pebbles. The children were then asked what would someone (who had not been shown) believe was in the box. They were then asked what God would believe. Three-year-olds answered pebbles in both

¹³⁰Ibid., 297.

¹³¹Ibid.; See M. Taylor, *Imaginary Companions and the Children Who Create Them* (New York: Oxford University Press, 1999).

¹³²Kelemen, "Are Children "Intuitive Theists"?: Reasoning About Purpose and Design in Nature." See R. Richert Justin Barrett, and A. Driesenga, "God's Beliefs Versus Mother's: The Development of Non-Human Agent Concepts," *Child Development* 72, no. 1 (2001).

cases: a true response in the case of God and a false response in the case of human beings. Interestingly, the older children understood that human beings would have a false belief and answer “crackers,” but even with this understanding of false beliefs in place, they answered “pebbles” for God, showing a distinction between the mental states of God (where representation is reality) and human beings (where representation and reality may be distinct). These findings suggest that around five-years of age children have the ability to distinguish between human beings and God when attributing mental states. Kelemen also found evidence for the presence of the third prerequisite—the understanding of artifacts and design:

[A]n underlying developmental pattern does emerge across all of these studies. With some reliability, the findings suggest that beginning sometime around the kindergarten period, children adopt a design-based teleological view of objects with increasing consistency. In light of this work, and earlier described research on children’s reasoning about nonnatural agent’s mental states, the proposal that children might be intuitive theists becomes increasingly viable.¹³³

Such findings, of course, have a significance that extends far beyond childhood.

Kelemen argues that this teleological stance is likely to remain the cognitive default: “If, as suggested here, the tendency to think in teleological quasi-artifact terms is a side effect of human mental design (and pan-cultural experience with artifacts) rather than socialization, it is likely to remain as a default explanatory strategy throughout life, even as other explanations are elaborated.”¹³⁴ Kelemen cites several factors that seem to support such a conclusion. First, reasoning about nature in “nonteleological physical-

¹³³Kelemen, “Are Children “Intuitive Theists”?: Reasoning About Purpose and Design in Nature,” 299.

¹³⁴Ibid.

reductionist” terms is a relatively recent development and “contemporary adults are still surprisingly bad at it.”¹³⁵ Studies also indicate that college age students will use teleological explanations outside of scientific contexts. Finally, research with scientifically uneducated Romanian Gypsy adults has found that they have promiscuous teleological intuitions much like scientifically naïve British and American elementary-school children.¹³⁶

Second-Order Theory of Mind and Symbolic Communication

We have been examining naturalistic sources of dualism, and it was argued that such dualism is relevant not only in understanding soul concepts but in understanding the genesis of God concepts as well. However, God concepts are not merely “big soul” concepts; other factors *alongside* dualism contribute to the formation of God concepts. Hyperactive agency detection leads us to see “faces in the clouds.” Theory of mind leads us to over-attribute purpose and design to the biological and non-biological natural world. But surely other notions are important as well. One important distinction between soul concepts and God concepts is epistemic. It is a truism that human beings have limited knowledge; this is especially true in the social sphere. Human beings seem to enjoy a kind of privileged access to their own mental states, whereas the mental states of other can only be inferred indirectly (*vis-à-vis* theory of mind). The limited knowledge of human beings stands in contrast to the omniscience characteristic of God concepts and other supernatural concepts. A central aspect of this omniscience is the notion that

¹³⁵Ibid.

¹³⁶Ibid., 300.

supernatural agents also enjoy privileged access to the mental states of human beings. As

Jesse Bering and Dominic Johnson write:

A central component of religious systems are concepts of supernatural agents that have privileged access to the self's mental states. Religious scholars would be hard pressed to find a religious system that does not have within its ranks some supernatural agent that, among other impressive facts, is envisioned as knowing rather than merely inferring from behavior the self's true intention.

Whereas deception is possible with human beings—one can withhold strategic information from them—it is not possible with God. God is a “full access strategic agent,” as Pascal Boyer puts it.

Now, the notion of privileged access goes hand in hand with the idea that God symbolically communicates through various positive or negative events. As Bering and Johnson write: “Moralizing gods can only find their way into large social groups insofar as individuals are capable of envisioning these gods as enforcing their morals through the occurrence of positive and negative events. A moralizing god who fails to ‘communicate’ with its followers would not be a very effective one.”¹³⁷

Attributing communicative intent is even more complex than agency detection and more complex than design attribution. Jesse Bering has convincingly argued that understanding events as symbolic communication requires having *second order* theory of mind in place, i.e., “knowing that God knows that I know/desire/etc. X.” “We argue, it is only with the appearance of second-order theory of mind that the child begins to see the natural events occurring in his life as meaningful, or more specifically, as symbolic and

¹³⁷Jesse M. Bering and Dominic D.P. Johnson, ““O Lord...You Perceive My Thoughts from Afar”: Recursiveness and the Evolution of Supernatural Agency,” *Journal of Cognition and Culture* 5, no. 1-2 (2005): 125.

declarative of an abstract intentional agent's desire to share social information with him."¹³⁸

Bering and his colleague Dorothy Parker tested this developmental claim on children aged 3-9. In this experiment, children played a forced choice game, where they had to guess which of two boxes contained a hidden ball. The children were to guess the location of the ball by placing their hand on the top of the box they believed contained the ball. The children were informed that an invisible agent (Princess Alice) would help them play by "telling them, somehow, when they chose the wrong box."¹³⁹ An experimenter would trigger an unexpected event (lights flashing on and off or a picture falling from the wall) and children's response to the event was coded. Only the oldest children reported declarative agentic responses—an unexpected event intentionally caused by an agent to communicate symbolic intent.¹⁴⁰ Middle aged children saw the unexpected events as caused by Princess Alice without communicative intent. The youngest children offered purely physical explanations.¹⁴¹

Bering's research indicates that the cognitive capacities that play a role in generating various religious representations like soul and God concepts develop at different stages. With intuitive dualism and afterlife beliefs this occurs quite early. This intuitive dualism likely plays an important role in the construction of God and spirit concepts, as both concepts involve the notion of "a person/mind without a body."

¹³⁸Ibid.

¹³⁹Bering and Parker, 255.

¹⁴⁰Ibid., 256-259.

¹⁴¹Ibid.

However, God and spirit concepts are more than simply persons without bodies; they are persons without bodies who both create and/or communicate. Entertaining such notions involves the capacity to see the natural world and items in the natural world as created for a purpose, and seeing certain events as symbolic communications. These features likely emerge at a somewhat later developmental stage as by-products of first and second order theory of mind.

Conclusion

The goal of this chapter was to provide a naturalistic explanation of the cross-cultural ubiquity of dualism. Such an explanation is part of the burden of proof for naturalists; they have to show not only that dualism is *false*; they also have to provide a compelling explanation of dualism. This chapter was an exercise in such explanation. It provided naturalistic explanations of dualism both on the levels of neuroscience and cognitive science.

In terms of neuroscience, it was shown that one major source of dualist intuitions concerns our folk phenomenology: this folk phenomenology generates both “out-of-brain” and “out-of-body” experiences. Such experiences are real and vivid enough, but can be explained in entirely naturalistic terms in light of the notions of a transparent phenomenal self model and bodily self-representation. In terms of cognitive science, it was shown that dualism emerges as an evolutionary by-product of other directly adaptive inference systems: folk physics, folk biology, and folk psychology. Dualism emerges as a consequence of the fact that we have two different inference systems for reasoning about social reality and reasoning about physical reality and from the fact that some mental

states are easier to imagine being without than other mental states. The fact that dualism comes so naturally to human beings also contributes to the genesis of other supernatural concepts like god concepts. Here dualism couples with another by-product of human cognition, the tendency to over-attribute agency and design. The promiscuous nature of agency detection and theory of mind lead us to see “faces in the clouds,” to see biological and non biological natural world as designed for a purpose, and to see events as symbolic communications.

The purpose of Chapters One and Chapter Two was to examine the implications of contemporary neuroscience, cognitive science, and philosophy of mind for our understanding of ourselves and our understanding of religion. Its fundamental argument is that the biggest implication of the contemporary science and philosophy of the mind-brain is that it calls into question mind-body dualism: the notion that a self can exist without a body, that there can in fact be Swinburnian “persons without bodies.” Chapter One was largely concerned with arguing for the truth of this position. However, contemporary neuroscience and cognitive science not only call into question mind-body dualism, they also have begun to offer compelling explanations as to why we almost inevitably think in dualist categories. Chapter Two argued that contemporary mind science is in fact well positioned to explain the natural sources of the dualist self. Having made this argument, the dissertation will address the implications of these for religion. This is the focus of Chapter Three.

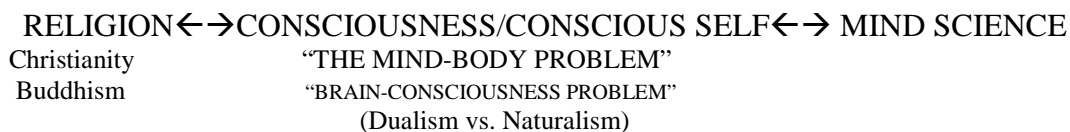
CHAPTER THREE
BIO-PSYCHOLOGICAL EXPLANATION AND THE TRUTH OF RELIGIOUS
DUALISM

Recap: Why Mind-Brain Science Matters for the Study of Religion

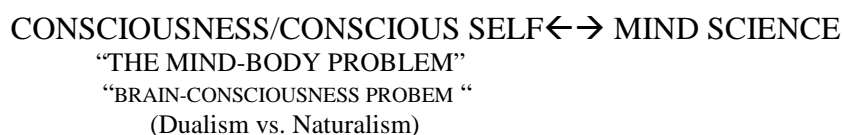
This dissertation began with the claim that scientific and philosophical work on consciousness and the conscious self is essential for the academic study of religion, particularly aspects of the academic study of religion involving the intersection of religion and science. Many religious traditions are rooted in strong forms of mind-body dualism, where consciousness and the conscious self are seen as distinct from brains and bodies, and capable of existing independently of them. Religious dualism is typically a two-fold dualism. Not only are human beings understood in a dualist manner, but all of reality is seen as grounded in some sort of immaterial mind or consciousness.

Chapters One and Two investigated the truth of such dualist claims about human persons and also explored how such dualist beliefs (including dualist religious beliefs) are generated. This chapter turns more explicitly to matters of religion. In particular, it is concerned to address the following question. Given the demise of dualism as a doctrine about human persons and the explanatory power of a naturalistic account of dualism, is any form of religious dualism a viable intellectual option today? This chapter argues that religious dualism is no longer a viable option. One seeking to integrate the claims of religion and science should thus look elsewhere.

The overall argument structure of the dissertation might be diagramed as follows:



Chapters One and Two focused primarily on *human minds*, specifically, the impact of the mind sciences on the contemporary mind-body problem or the problem of consciousness.



Chapters One and Two argued that contemporary mind science calls into question dualist claims about consciousness while simultaneously explaining their origin in human thought. Chapter One presented the debate between dualism and naturalism as a debate about the adequacy of bio-psychological explanation of consciousness, and argued that principled dualist and mysterian objections to a scientific, bio-psychological explanation of consciousness are simply not compelling. Such claims are rooted in inadequate deductive criteria of explanation and reduction, and, as such, fail to appreciate the resources a naturalistic, bio-psychological framework has for taking the subjectivity of consciousness seriously. The mind-body problem, or, more specifically, the brain-consciousness problem, was resolved in favor of a neurophilosophical, inflationist form of naturalism. Chapter Two then argued that the contemporary mind sciences not only call dualism into question, they also help explain the genesis of robust dualist intuitions in human thought.

Chapter One and Chapter Two also corrected methodological misconceptions that naturalism is often saddled with. Naturalism does not entail a deflationist reduction of phenomena like phenomenal consciousness, nor does it entail a collapsing of all higher-level disciplines into biology, much less physics. What naturalism does rule out are claims about the ontological independence of consciousness and claims about the methodological independence of higher level disciplines such as the humanities and social sciences. Thus we have addressed the debate between dualism and naturalism concerning human minds and in the process corrected some methodological worries about naturalism.

If these arguments have been successful, there is every reason to be skeptical about dualism as a doctrine about human persons. What sort of implications does this have for religion? This is the central issue in the current chapter of this dissertation. We are now moving from questions about mind science and the mind-body problem to explicitly address matters of religion.

RELIGION	←→	CONSCIOUSNESS/CONSCIOUS SELF	←→	MIND SCIENCE
Ch. 3		“THE MIND-BODY PROBLEM”		Ch. 1-2
		“BRAIN-CONSCIOUSNESS PROBLEM”		

Religious dualism was introduced in the beginning of the first chapter. Chapters One and Two then offered a philosophical and scientific critique and explanation of dualism. The truth of dualism as a philosophical thesis about consciousness was called into question in Chapter One, and Chapter Two offered a naturalistic explanation of the ubiquity of mind-dualism. This Chapter explicitly explores the implications of the

conclusions of Chapter One and Chapter Two for religious dualist claims with the Christian and Buddhist traditions.

In the first section, I establish the intimate link between dualism and theism through an analysis of classical theism. I argue that classical theists are right that the rejection of dualism has some fairly significant theological fallout, but they are wrong in thinking that substance dualism can be saved as a doctrine about human persons. Classical dualists rely on a priori philosophical arguments that were shown to be unsuccessful, and their work fails to make any significant contact with actual work in the mind sciences. The remainder of the chapter is devoted to an analysis of attempts to integrate religion and the mind sciences.

Many Christian theologians interested in “religion and science” deny dualism as a doctrine about human persons while seeking to preserve a dualist doctrine of God. I consider what is perhaps the most sophisticated version of this position—Philip Clayton’s synthesis of emergentism and theism. Claims about immaterial minds and consciousness surface not only in the dialogue between Christianity and science but in the Buddhist-mind science dialogue as well. I conclude the chapter with a consideration B. Alan Wallace’s attempt to integrate Buddhism and mind science through his “contemplative science” project.

Dualism and Classical Theism

There are very significant links between dualism and theistic belief. The aim of this section is to make this link explicit. Theistic religious traditions frequently endorse a dualist understanding of human nature. In particular, within Christian theism, a dualist

understanding of human nature certainly figures prominently. This is perhaps most apparent in matters of eschatology and afterlife beliefs. Contemporary theologians like to remind people that Christian teaching involves resurrection of the body rather than immortality of the soul, but this is surely a false dichotomy.¹ Perhaps the majority of the ecumenical Christian tradition has held a view that combines the notions of the immortality of the soul and resurrection of the body. As Charles Taliaferro notes:

In general, I think the Christian understanding of the afterlife is best served by assuming a dualist view of human nature. It is widely believed in the Christian tradition that persons exist after the destruction of their body and that, at the resurrection, they will receive either a new embodiment or an embodiment that is constituted by their reconfigured, transformed earlier body. The personal life between death and resurrection is often thought of as either disembodied or as some intermediary embodiment...These scenarios all assume some form of dualism, for if materialism is true and the person is her body, then the annihilation of the body entails the annihilation of the person.²

John Cooper is forthright about the consequences of a rejection of dualism for this aspect of Christian belief:

Many in the academic community have taken a clear position on the body-soul question which they continue to assert with conviction. And if what they are saying is true, then two disturbing conclusions immediately follow. First, a doctrine affirmed by most of the Christian Church from the beginning is false. A second consequence is more personal and existential—what millions of Christians believe will happen to them when they die is also a delusion...If souls are not the sort of thing which can be broken loose from bodies, then we do not actually exist between death and

¹The contrast between immortality of the soul and resurrection of the body was first made explicit by the German biblical scholar, Oscar Cullman in his 1956 article, “Immortality of the Soul or Resurrection of the Dead.” Since then, a number of scholars have called into question the sharp distinction between Greco-Roman and Hebraic thought found in Cullman’s writings. Nonetheless, such distinctions enjoy a great deal of prominence in theological discussions.

²Charles Taliaferro, *Consciousness and the Mind of God* (Cambridge, UK: Cambridge University Press, 1994), 245.

resurrection, either with Christ or somewhere else, either consciously or unconsciously.³

Much more is at stake with the issue of dualism than matters of eschatology.

There exists an intimate link between dualist notion of human persons and the conception of God within classical theism. Theism, in fact, is a form of dualism and also functions as the explanation of dualism regarding human persons. Dualism as a doctrine about human nature thus has very significant bearing on fundamental issues concerning the nature and existence of God. Maurice Schouten describes this link well:

In classical theism, the (personal) properties we typically ascribe to our fellow human beings and to ourselves are projected onto God. Thus, the view of God is modeled after the dualist image of humanity. God is seen as an incorporeal personlike being with perfectly rational and conscious thought who intervenes in the material world on the basis of knowledge, and all of this without bodily features. God listens, hears, sees, believes, knows, wants, and so forth—but all without ears, eyes, a mouth, or a cerebral cortex. And this image comports nicely with the traditional picture of ourselves according to which the spiritual part of the human agent directs its bodily part and acts out into the material world he or she is surrounded by.⁴

Theism thus depends on dualism to secure a notion of immaterial personhood, a notion of God as a “person without a body.” As Charles Taliaferro puts it,

In western tradition God is typically conceived of as an immaterial, spiritual being. God is a nonphysical, powerful, intentional agent present throughout the cosmos without being identical with it or with some material object in it...If a materialist view of the cosmos, according to which *all* of reality is fundamentally physical, is correct, then the traditional understanding of God is incorrect.⁵

³John Cooper, *Body, Soul, and Life Everlasting: Biblical Anthropology and the Monism-Dualism Debate* (Grand Rapids: Eerdmans, 1989), 1.

⁴Maurice K.D. Schouten, "Theism, Dualism, and the Scientific Image of Humanity," *Zygon* 36, no. 4 (1991), 683.

⁵Taliaferro, 2.

Taliaferro is, of course, aware that God is not a super soul or as he puts it a “super *homo sapiens*,” but he also has the awareness that theism involves making analogical predications based on human cognition and that such predications cannot get off the ground given a naturalist view of the human person:

In the great monotheistic religious traditions, God is not, of course, depicted with precisely the same attributes as created persons; God is not a super *homo sapiens*. And at important junctures in the development of these religions, theists have gone to great lengths to insist that God differs from us in that God is eternal; while we are temporal; God is not made up of parts, we are; God has no sense organs, we do; God creates galaxies from nothing, we only shape things that already exist; God is without origin and incorruptible, we are neither. In the central conviction that God exists necessarily, while we exist only contingently it might appear that the God of classical theism resembles abstract propositions like “ $7+5=12$ ” which are necessarily true far more than God resembles us. But God is also believed to know about the world, to be an all-good agent, to reshape things which already exist, to love the cosmos. Theists will diverge on how to understand these features as ascribed to God, but it is evident that these ascriptions make up central items in the folk-psychological framework that is under attack in much contemporary philosophy. Theists refine terms like “intelligence,” “agency,” and “emotions” as applied to God, but surely the theistic enterprise requires that we have some conceptual handle on our own intelligence, agency, and emotions before religious refining can take place. And if...materialism...holds sway, it is difficult to see how such theistic refining can even get started.⁶

A dualist doctrine of the human person is intimately related not just to questions concerning the *nature* of God but also to questions concerning God’s existence. A dualist doctrine of the human person must explain how an immaterial human mind or soul could originate and then become conjoined with a particular body. Theism has thus also served as the explanation for a dualist notion of human persons; divine consciousness explains human consciousness. This finds expression in notions of God as the direct creator of the

⁶Ibid., 5-6.

human soul and in attempts to argue from the fact of human consciousness to the existence of God.

Roman Catholic intellectuals frequently like to point out that Catholicism is completely compatible with the theory of evolution. They frequently point with approval to Pope John Paul II's 1996 address to the Pontifical Academy of Sciences, "Truth Cannot Contradict Truth." In this address John Paul goes beyond Pope Pius XII's statement of compatibility in his 1950 encyclical *Humani Generis*, and proclaims that a half century of evidence leads to the conclusion that evolution is "more than a hypothesis":

Today, almost half a century after the publication of the encyclical [*Humani Generis*], new knowledge has led to the recognition of the theory of evolution as more than a hypothesis. It is indeed remarkable that this theory has been progressively accepted by researchers, following a series of discoveries in various fields of knowledge. The convergence, neither sought nor fabricated, of the results of work that was conducted independently is in itself a significant argument in favor of this theory.⁷

But, it is important to note that the former pope is speaking of the *evolution of the body*.

The mental life of human beings, i.e. the soul, must be understood as directly created by

God:

The Church's magisterium is directly concerned with the question of evolution, for it involves the conception of man: Revelation teaches us that he was created in the image and likeness of God (cf. Gn 1:27-29). The conciliar constitution *Gaudium et Spes* has magnificently explained this doctrine, which is pivotal to Christian thought. It recalled that man is "the only creature on earth that God has wanted for its own sake" (No. 24). In other terms, the human individual cannot be subordinated as a pure means or a pure instrument, either to the species or to society; he has value *per se*. He is a person. With his intellect and his will, he is capable of forming

⁷John Paul II, "Truth Cannot Contradict Truth," (1996). Retrieved from: http://www.newadvent.org/library/docs_jp02tc.htm (accessed 10/10/09).

a relationship of communion, solidarity and self-giving with his peers. St. Thomas observes that man's likeness to God resides especially in his speculative intellect, for his relationship with the object of his knowledge resembles God's relationship with what he has created (*Summa Theologica* I-II:3:5, ad 1). But even more, man is called to enter into a relationship of knowledge and love with God himself, a relationship which will find its complete fulfillment beyond time, in eternity. All the depth and grandeur of this vocation are revealed to us in the mystery of the risen Christ (cf. *Gaudium et Spes*, 22). It is by virtue of his spiritual soul that the whole person possesses such a dignity even in his body. Pius XII stressed this essential point: If the human body takes its origin from pre-existent living matter, the spiritual soul is immediately created by God. Consequently, theories of evolution which, in accordance with the philosophies inspiring them, consider the spirit as emerging from the forces of living matter or as a mere epiphenomenon of this matter, are incompatible with the truth about man. Nor are they able to ground the dignity of the person.⁸

As claims about God as creator of the cosmos find expression in cosmological arguments, claims about God as the direct creator of the soul find expression in arguments from consciousness. Such arguments date back at least to John Locke and have recently been revived by Richard Swinburne, Robert Adams, and J.P. Moreland. Swinburne, for example, argues that while science cannot explain the relationship between the objective brains and subjective selves, theism can:

But theism can provide an explanation of these things. God, being omnipotent, is able to join souls to bodies. He can cause there to be the particular brain event-mental event connections which there are. He can do this by causing molecules when formed into brains to have powers to produce mental events in souls to which they are connected, and the liabilities to execute the purposes of such connected souls (new powers and liabilities not deriving from the ordinary ones which chemistry analyses). And he can make the souls in the first place and choose to which brain (and so body), each soul is to be connected when foetal brain events require a soul to be connected to the brain.⁹

⁸Ibid.

⁹Richard Swinburne, *Is There a God?* (Oxford Oxford University Press, 1996), 90.

J.P. Moreland nicely outlines the premises of the “argument from consciousness”

taken as a straightforward deductive argument:

- (1) Mental events are genuine non-physical mental entities that exist.
- (2) Specific mental event types are regularly correlated with specific physical event types.
- (3) There is an explanation for these correlations.
- (4) Personal explanation is different from natural scientific explanation
- (5) The explanation for these correlations is either a personal or natural scientific explanation.
- (6) The explanation is not a natural scientific one.
- (7) Therefore, the explanation is a personal one.
- (8) If the explanation is personal, then it is theistic.
- (9) Therefore, explanation is theistic.¹⁰

All of these arguments are premised upon very strong forms of substance dualism, and, as such, rest on very strong versions of the Correlation Claim introduced in Chapter One: the claim that the relationship of mind and brain can only be understood in terms of correlation and not identity.

All dualists hold the Correlation Claim and thus deny mind-brain identity, but body-soul dualists deny even the dependence/supervenience of the human mind/soul on the brain (indeed they must in order to claim God as the direct creator of the soul and preserve the argument from consciousness). Classical theists who are also substance dualists thus hold not only the Correlation Claim but the more robust No Neural Correlates Claim for both the soul and for God: Consciousness or the conscious self is capable of existing without any neural or physical correlates (e.g., the post-mortem soul, God, substrate consciousness).

¹⁰J.P. Moreland, *Consciousness and the Existence of God* (London: Routledge, 2008), 37.

In short, classical religious dualists recognize the intimate link between substance dualism and theism and recognize how even mild versions of mind-brain dependency/supervenience compromise theism. Philip Clayton, a theologian who rejects substance dualism in favor of an emergentist doctrine of the human person, also expresses this worry:

[T]he case for emergent mental causation is not by itself a case for the existence of God, divine action, an eternal soul, or life after death; it is not directly a theological conclusion at all. Indeed, in some ways it might seem to be an *anti*-theological conclusion, because it understands mental phenomena to be “of a piece” with physical phenomena, and because the supervenience relationship asserts a dependence of the mental life on its physical basis—indeed a high correlation between physical causes and mental effects—which is on the surface inconsistent with many parts of Christian teaching.¹¹

What is significant to note is that the arguments of Chapter One called into question even *mild* versions of the correlation claim, as expressed in emergentist and property dualist positions. If property dualism is untenable, substance dualism is a fortiori untenable. The strongest arguments for property dualism are knowledge arguments and modal arguments that attempt to show that conscious mental properties cannot be physical properties and that the mind-brain relation can only be understood in terms of correlation and not identity.

Chapter One considered in some detail knowledge and modal arguments proposed by Thomas Nagel, Saul Kripke, Joseph Levine, Frank Jackson, and David Chalmers. The flaw of such arguments rests on a failure to understand the incompleteness of our

¹¹Philip Clayton, "Neuroscience, the Person, and God: An Emergentist Account," in *Neuroscience and the Person: Scientific Perspectives on Divine Action*, ed. Nancey Murphy Robert John Russell, Theo C. Meyering, Michael A. Arbib (Vatican City State: Vatican Observatory Publications, 2002), 206.

concepts linking consciousness and the brain, a failure to appreciate the resources naturalism/physicalism has for accommodating subjectivity, and from inadequate deductive criteria of explanation. Chapter One went on to defend an inflationist naturalistic account of consciousness in the philosophy of mind, and explanatory pluralism and mechanistic explanation in the philosophy of psychology and biology. It was then shown how the hard problem of consciousness could be pursued in an empirically tractable manner. Chapter Two then offered a fully naturalistic account of dualist “intuitions of distinctness.”

A number of neurological and psychological factors contribute to our dualist intuitions. We do not directly sense the brain; certain mental states are more intimately connected with the body than others. Brain states are also unique in nature in that they are both perspectival and transparent. Brain states are unique in that they can be experienced from both a first person and a third person perspective: subjectively and objectively. Subjective physical facts are just objective physical facts experienced by a particular organism. I can experience the phenomenal level in my brain from a first person and third person perspective. Such first person facts do not seem very brainy, due to the transparency of consciousness, the fact that “the brain is not about itself”;¹² it is a form of “virtual reality” generated by the brain that *seems to* take us out of the brain and place us in immediate contact with reality. For these reasons, dualist thinking comes fairly easily to us, but it ultimately is not true.

¹²See Todd Feinberg, *Altered Egos: How the Brain Creates the Self* (Oxford: Oxford University Press, 2001).

Is this a *definitive* disproof of dualism? No, but arguments based on consciousness are widely seen as the strongest weapons in the dualist arsenal. But dualist arguments about correlation fail to work; moreover dualism fails to cohere well with the physics, biology, and neuroscience constraints discussed earlier. Dramatic alterations in conscious phenomenology and the phenomenal self can be attributed to specific regions in the brain, the dualist owes us an explanation of how the self can be so fundamentally affected by minor alterations in the brain and yet remain ontologically independent of it. The question of *how* the conscious self evolved remains a vexed question, but researchers assume *that* consciousness evolved and have offered educated guesses as to its selective function. The dualist owes us an answer as to how these approaches are inadequate and needs to provide a better constructive alternative. If the soul is directly created by God, when exactly did this process occur—with australopithecines? With the appearance of the genus *homo*? When indeed.

Even more pressing is the question of how dualists might account for mental causation without violating conservation laws. Theistic substance dualists—apart from occasionalists—also typically want to affirm causal interaction between the mental and the physical. It seems obvious that consciousness mental occurrences have physical effects in the world, but conservation laws and the principle of the causal closure of the physical place great constraints on how one thinks about such causation. Interactionist dualism was jettisoned by Descartes' contemporaries Malebranche and Leibniz precisely for this reason, and the most prominent defenders of dualism today also reject interactionist dualism. Note that if a version of identity theory like the one defended in

Chapter One is true, the problem of mental causation is saved as a species of physical causation. Naturalism can fully account for the subjectivity of consciousness and offer the most plausible account of mental causation.

The conclusion of all of this is that body-soul dualism is simply no longer a compelling position, but substance dualists are fundamentally correct in noting the significant link between dualism and theism. Dualism and theism are so intimately linked that it would seem that the collapse of mind-body dualism has rather dire consequences for religion, particularly with regard to claims about God as a “person without a body.” But might it be possible to reject dualism as a doctrine about human persons while defending a dualist doctrine of God? Such a position has in fact become increasingly common, especially with Christian philosophers and theologians concerned with engaging the mind sciences and analytic philosophy of mind. Those defending such a position are placed in the awkward position of affirming a downward dependence of human minds on brain but denying that transcendent mind is downwardly dependent too. An awkward position, however, is not an impossible position and the next section considers a sophisticated attempt to combine an emergentist doctrine of the human person with a dualist doctrine of God.

Emergentism and Theism: Philip Clayton

The previous section discussed the intimate connection between dualism, theism, and key Christian beliefs. The point of this section was to note that the demise of dualism in the philosophy of mind has fairly significant consequences for Christian theology as traditionally conceived. It is not just that one kind of argument for the existence of God

has been cut off. Rather, the more disturbing question is raised as to why dualism should be pursued at the theological level if it makes no sense at the level of human beings. The notion of God as a “person without a body” becomes much less intelligible when body-soul dualism about human beings collapses, and when one can offer naturalistic accounts of the mechanisms that lead human beings to think in dualistic terms. The Christian theologian is faced with quite a dilemma. One can either continue to affirm body-soul dualism and traditional afterlife beliefs at the high cost of rejecting a great deal of scientific and philosophical evidence, or one can accept the philosophical and scientific evidence and reject some of the most foundational truth claims in the Christian tradition.

Many theologians simply choose to ignore the philosophical challenges posed by the mind sciences or minimize the significance of these challenges to theology. Philip Clayton is a notable exception. Clayton has extensively engaged issues concerning religion and the mind sciences and is well aware of the pressures that the mind sciences put on theology. He seeks a middle ground between dualist and physicalist positions through the notion of emergence. He then uses this emergentist account of the human person to argue for a form of theistic *dualism*. This section considers Clayton’s approach to theology and neuroscience, and then turns to Clayton’s emergentist understanding of the human person and his arguments for theological dualism. It is argued that there are reasons to be skeptical of Clayton’s attempt at integrating theology and the mind sciences, based both on his notion of emergence and his attempt to move from emergence to dualist transcendence.

Clayton's Account of the Relationship Between Theology and Neuroscience

Clayton senses the pressures that the physicalist assumptions of the neurosciences put on theology; however, for Clayton, the task of theology is not to critique these assumptions or to articulate a proposal for a soul-based neuroscience. The theological task rather involves seeing how far one can go in embracing these assumptions without denying factors necessary for the viability of religious belief. “[T]he guiding question in the dialogue between theology and the neurosciences is, how far can a position go in the direction of the physicalist assumptions that are basic to the empirical study of the brain without denying (or implicitly rejecting) factors necessary for the viability of religious belief?”¹³ Clayton argues that certain minimal conditions of personhood are necessary if a realist version of theism is to remain a viable option:

It includes, at least for traditional theologians, not only the existence of at least one purely spiritual being—hence the possibility of disembodied agency—but also the notions of will and of freedom, which come in both finite and infinite flavors. With will, so understood, comes consciousness: Christians conceive God as a conscious agent, an agent enough like human agents that the predicate “person” can also be attributed, if only in an analogous fashion, to the divine. On this view, humans and God are also *moral* agents; persons exercise their agency in light of real obligations to other persons (indeed, to the world as a whole) and to God. Finally, these agents are social agents. Religious notions of *community* emphasize a union among humans in light of the divine presence and the covenant which makes of us “one.”¹⁴

Clayton is forthright about the pressures that the mind sciences put on these conditions, especially on the notion of disembodied mind or agency:

¹³Philip Clayton, "Neuroscience, the Person, and God: An Emergentist Account," 184.

¹⁴*Ibid.*, 182.

Part of the problem is that we are no longer sure what to make of the notions of mind or spirit. The metaphysical resources of the Western tradition—the conceptual world of *ruach*, *pneuma*, *spiritus*, *Geist*—are difficult to reconcile with the attitude and results of contemporary science. One can of course still assert that ‘God is Spirit, infinite and perfect in his being and perfections,’ as the Westminster Confession has it, one can affirm that humans are made ‘in the image of God’ (*imago dei*); and one can conclude that each human therefore possesses a God-like spirit, as the Pope recently reaffirmed in his statement on evolution. But whereas this view once accorded nicely with the natural science (natural philosophy) of a previous era, it stands in deep tension with the approaches and the results of the science of our own.¹⁵

For Clayton, it is the notion of emergence that allows for a critical mediation between theology and the neurosciences:

That is where the emergence argument comes in. If successful, this argument represents a *tertium quid* between physicalist treatments of mind, which leave no place for talk of spirit, and dualist treatments, which simply assume (in my view, too easily) the continuing validity and usefulness of such language.¹⁶

Clayton’s Account of Emergentism

In his monograph, *Mind and Emergence*, Clayton defends an emergentist account of the natural world and the human person in particular. He then uses that emergentist account of the human person to argue for a theistic *dualism*. Clayton understands his emergentist position as a *via media* between physicalism and dualism.

It is widely but falsely held that there are only two major ways to interpret the world: in a physicalist or in a dualist fashion.... The evolutionary perspective has fatally undercut both sides of the once regnant either/or: physicalism, with its tendency to stress the sufficiency of physics, and dualism, with its tendency to pull mind out the evolutionary account altogether.... I argue that *emergence* is the philosophical position—more

¹⁵Clayton, *Mind and Emergence: From Quantum to Consciousness*, 185-186.

¹⁶*Ibid.*, 186.

accurately, the philosophical elaboration of a series of scientific results—that best expresses the philosophical import of evolutionary theory.¹⁷

This positioning of the issues is important to keep in mind, and Clayton’s treatment tends to focus on three positions:

PHYSICALISM—EMERGENTISM—DUALISM

A key issue in evaluating Clayton’s work is the meaning of the notion of “emergence” and its relationship to terms like “dualism,” “reductionism,” and “physicalism.” In analyzing Clayton’s work, it is important to first distinguish between three senses of emergence.¹⁸ I will call these (1) trivial emergence, (2) weak emergence, and (3) strong emergence. (1) Trivial emergence or non-emergence is essentially the notion of aggregativity: systems acquire new causal powers based on an aggregation of their parts. For example, a five-pound hunk of iron has different causal powers than a single iron atom, but the causal powers of the hunk are aggregative. (2) Weak emergence captures the fact that certain systems are not mere aggregates of their parts; organizational interactions allow the system to do things that the parts cannot do. Weak emergence is essentially the notion of emergence as non-aggregativity. Weak emergence can also refer to the fact that certain systems are so complex that their behavior cannot be predicted in practice from what is known about the organization of its components. Weak emergence in this sense refers to epistemological limitations. In contrast to these notions of emergence, the notion of strong emergence or radical property emergence holds that

¹⁷Ibid., 1.

¹⁸My taxonomy is indebted to discussion of emergence in the work of William Wimsatt, Carl Craver, Carl Gillett, and Terrence Deacon.

systems give rise to *sui generis* properties that cannot be explained in terms of their component parts plus organization.

Given these various understandings of “emergence,” I’m not convinced that emergence represents a genuine *tertium quid* between physicalism and dualism. Strong emergence is actually a form of property dualism, while weaker notions of emergence—emergence as non-aggregativity and epistemological emergence—can be incorporated into a physicalist, and even a reductive physicalist framework. As William Wimsatt has observed:

An opposition between reduction and emergence forces people to take sides along an axis missing some of the most revealing cuts of the issue. One can be a reductionist and an emergentist too, with a proper understanding of these notions. Misunderstandings engender opposition to reductionism and make emergence unnecessarily mysterious.¹⁹

I think Wimsatt is fundamentally right on this issue. It is difficult to overestimate the amount of mischief faulty understandings of reductionism have caused in philosophy of mind and philosophy of science. The notion of reduction that has perhaps generated the most mischief is an understanding indebted to positivist philosophy of science where reduction is understood as a logical relationship between theories that bottom out at a privileged level of explanation (typically that of microphysics). Two aspects of this understanding of reduction are especially important: (1) reliance on deductive criteria of explanation, (2) emphasis on the notion that scientific explanations bottom out in a privileged set of entities or causal relations. What is significant to note is that

¹⁹William Wimsatt, “Emergence as Non-Aggregativity and the Biases of Reductionisms,” in *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality* (Cambridge, MA: Harvard University Press, 2007), 274.

contemporary philosophers of psychology, neuroscience, and biology have jettisoned both Deductive-Nomological models of explanation and fundamentalist²⁰ reductionism in favor of mechanistic models emphasizing a plurality of levels of explanation. Advocates of mechanistic explanation are only anti-reductionist in the sense that they repudiate the notion of reduction as deduction and the privileging of a single level of explanation (fundamentalist reduction). As William Wimsatt notes, under a mechanistic model of explanation, “A reductive explanation of a behavior or property of a system is one that shows it to be mechanistically explicable in terms of the properties of and interactions among the parts of the system.”²¹ This entire framework puts the issue of reductionism in a very different light. As William Bechtel observes:

[M]ost neuroscience explanations do not take the form of D-N explanations in which phenomena are derived from laws, but rather are models of mechanisms. This casts a different light on the issue of reduction. Models of mechanisms are inherently reductionist: each proposed mechanism is designed to show how a phenomenon ascribed to a system is due to its constituent parts and their interaction. On the other hand, reduction no longer threatens the autonomy of the higher-level science: the higher-level characterizes the interaction of processes, the lower-level accounts for the performance of individual processes.²²

The notion of “interaction of parts” mentioned both in the quotes by Wimsatt and Bechtel is central, for mechanisms are non-aggregates. If one understands emergence as non-aggrativity one can embrace a kind of emergence while remaining within a

²⁰By “fundamentalist” reductionism, I mean the doctrine that scientific explanations bottom at some privileged level. I am not equating reductionists with *religious* fundamentalists, an inane rhetorical move favored by a surprising number of theologians working at the intersection of religion and science.

²¹Wimsatt, 275.

²²Pete Mandik William Bechtel, and Jennifer Mundale, "Philosophy Meets the Neurosciences " in *Philosophy and the Neurosciences: A Reader*, ed. Pete Mandik William Bechtel, Jennifer Mundale, Robert Stufflebean (Oxford: Blackwell, 2001), 37.

physicalist and reductionist framework. One can also accept a form of epistemological emergence. None of these weak forms of emergence removes one at all from a physicalist—or even reductive physicalist-- framework. What is ruled out are strong versions of property emergence. Philosopher of neuroscience Carl Craver makes these distinctions well:

It is important to keep several different senses of the term “emergence” distinct. Some philosophers and scientists use the term “emergence” to describe properties of wholes that are not simple sums of the properties of components. Mechanisms are nonaggregates, and so they are emergent in this weak sense. Mechanisms require the organization of components in cooperative and inhibitory interactions that allow mechanisms to do things that the parts themselves cannot do. Other philosophers and scientists use the term “emergence” to mean that it is not possible to predict the behavior of a mechanism as a whole from what is known about the organization of its components. This is sometimes called “epistemic emergence” Some mechanisms have so many parts and such reticulate organization that our limited cognitive and computational powers prevent us from making such predictions. Such mechanisms are so sensitive to undetectable variations in input or background conditions that their behavior is unpredictable in practice. Behaviors of mechanisms are sometimes emergent in this epistemic sense. However, one who insists that there is no explanation for a non-relational property of the whole in terms of the properties of its component parts-plus-organization advocates a spooky form of emergence....The ability of organization to elicit novel causal powers (that is, nonaggregative behaviors and properties) is unmysterious both in scientific common sense and common sense proper. Appeal to strong or spooky emergence, on the other hand, justifiably arouses suspicion²³

The kind of emergence advocated by Clayton and fellow theistic emergentists William Hasker and Timothy O'Connor is a very robust form of strong emergence. Clayton is right to repudiate fundamentalist versions of reductionism and physicalism,

²³Carl F. Craver, *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience* (Oxford and New York: Oxford University Press, 2007), 216-217.

but his own work remains essentially crypto-dualist. Clayton clearly repudiates body-soul or substance dualism, but the version of emergence he defends is essentially a form of property dualism. In fact, it is a very strong form of interactionist property dualism. Clayton tends to eschew the term “property dualism” and opts instead for terms like “property pluralism” and “emergentist monism”; however, Clayton certainly sounds like a property dualist when he states: “The balance that we seek conceives mind as a type of property that emerges from the brain, which though different from remains continually dependent on its subvenient base (hence the term emergentist supervenience).”²⁴ Clayton wants to argue that strong property emergence is a feature of the cosmos “from quantum to consciousness,” and this is precisely why he prefers the term “property pluralism” to “property dualism”:

Clearly the position defended here is not a version of substance dualism; there has been no suggestion of mental substances intervening in the physical order. But is it a variant of *property* dualism, the view that, even if there is only one kind of substance, it has two fundamentally different kinds of properties?

Such a criticism rests on a misunderstanding. I have not portrayed a world divided into two distinct types of qualities, but rather a world with a vast array of different types of properties. Though there is no justification for the “dualism” label, the theory could fairly be called *property pluralism*, since it countenances a wide range of properties depending on their position in the complexity hierarchy.²⁵

Clayton cites a number of natural science examples to make his case for strong emergence, but the question can be raised as to whether the numerous examples Clayton cites are really examples of strong emergence or merely forms of weak emergence

²⁴Clayton, *Mind and Emergence: From Quantum to Consciousness*, 128.

²⁵Clayton, "Neuroscience, the Person, and God: An Emergentist Account," 212.

discussed above. There is certainly a danger of equivocation in bringing together a wide variety of examples of “emergence” from the natural sciences. Michael Silberstein makes this point well:

Clayton argues from enumerative induction to the conclusion of nomological [strong] emergence. Cases from the natural sciences that he canvasses include conductivity, chaotic hydrodynamics, Raleigh-Benard convection, autocatalysis, self-organization of various sorts such as the formation of snowflakes, Belousov-Zhabotinsky reactions, finite cellular automata, evolution, neural networks, ant colonies, the quantum Hall effect, quantum decoherence, and the Pauli Exclusion Principle. The only thing these myriad case studies have in common is that they have all been tagged ‘emergent phenomena’ at one point or another in the literature; but the important question is what kind of emergence they represent. Unfortunately, none of these cases is an obvious example for nomological or radical mereological emergence, and that is what Clayton needs to contradict causal closure and physicalism.²⁶

Clayton himself recognizes this ambiguity, but in his treatment of mind he argues that this ambiguity disappears:

In the previous chapter we found that some of the biological cases stand on the boundary between weak and strong emergence. I argued that the strong interpretation does better justice to biology as a whole, given that emergent systems are not just aggregates of microphysical states but cells and organisms—the agents that populate the biosphere and that serve as individualized objects of study for many biologists. Still, I had that at least some of the scientific cases could be read either way. In the case of mental phenomena, I will now argue, the ambiguity disappears: one cannot make sense of mental causation except from the standpoint of strong emergence. If the strong emergence interpretation of mental causes is not correct, one should be an epiphenomenalist about mind, that is, one should hold that mind has no effect on the world. To the extent that one thinks that epiphenomenalism is a conclusion to be avoided, to that extent one has reason for endorsing strong emergence.²⁷

²⁶Michael Silberstein, "Emergence, Theology, and the Manifest Image," in *The Oxford Handbook of Religion and Science* ed. Philip Clayton (Oxford: Oxford University Press, 2006), 791.

²⁷Clayton, *Mind and Emergence: From Quantum to Consciousness*, 108.

Clayton argues that both consciousness and mental causation demand a strong emergentist reading. Interestingly, Clayton’s discussion of consciousness and mental causation fails to make contact with actual scientific work on consciousness and agency. Clayton briefly cites nine examples of significant research on consciousness and agency, but ultimately uses these examples to argue for a version of the Correlation Claim, despite the fact that many of the scientists cited by Clayton—Francis Crick, Christoph Koch, Gerald Edelman, Giulio Tononi, Bernard Baars, and Wolf Singer—would contest the Correlation Claim. “Research into the neural correlates of consciousness—one of the most fruitful areas in the study of consciousness today—can offer no more than its name promises. At most one will be able to establish a series of *correlations* between brain states and phenomenal experiences as reported by subjects.”²⁸ In terms of consciousness, Clayton appeals to Frank Jackson and David Chalmers in defending his emergentist account of the human mind.

Chalmers’s own answer to the problem of consciousness...whatever other inadequacies it may have—does seek to explain what is different about the experiential states that persons have. What he elsewhere calls ‘naturalistic dualism’ is the right *sort* of answer to the hard problem.²⁹

Elsewhere, Clayton appeals to multiple realizability to argue against a type identity understanding of mind:

The fact of multiple realizability weakens the claim that mental properties are really of the same type as physical properties, a ‘type-identity’ theories hold. Thus pain is a different type of property than the chemical properties of a given neuronal synapse.³⁰

²⁸Ibid.

²⁹Ibid., 123.

³⁰Ibid., 126.

Somewhat incredibly, Clayton merely cites these canonical anti-brain arguments in the philosophy of mind and his work makes no contact with the fairly substantial critical literature that these arguments have generated. As Chapter One already discussed, arguments from consciousness and multiple realizability do not succeed. They are not arguments about the current status of science; they are principled arguments against the explanatory significance of neuroscience. Chapter One has given us every reason to be skeptical of anti-naturalist arguments based on both consciousness and multiple realizability.

Clayton's account of mental causation runs into even more serious difficulties. Clayton wants to save mental causation and avoid epiphenomenalism. He claims that his version of strong emergence is the only alternative that avoids epiphenomenalism. Somewhat oddly, Clayton defends the Jackson-Chalmers line on consciousness, but Jackson and Chalmers both recognize that their positions entail or come close to a form of epiphenomenalism. Jackson's first article containing the Mary thought experiment is significantly entitled "epiphenominal qualia." Chalmers also is a dualist only about qualia, and explicitly rejects interactionist forms of property dualism when he writes: "Interactionist dualism requires that physics will turn out to have gaps that can be filled by the action of a nonphysical mind. Current evidence suggests that this is unlikely."³¹

The view that only strong emergence is the only way one can account for mental causation is surely an idiosyncratic view. Indeed, philosophers such as David Papineau

³¹David J. Chalmers, *The Conscious Mind: In Search of a Fundamental Theory* (Oxford: Oxford University Press, 1996), 162.

and Jaegwon Kim have argued that given the causal closure of the physical—it is physicalism that can account for mental causation and jettison epiphenomenalism. Epiphenomenalism is a form of dualism not materialism. The naturalist can regard consciousness as robustly causal—indeed it would be odd if it were not causal. As David Papineau has observed, “If epiphenomenalism were true, then the relation between mind and brain would be like nothing else in nature. After all, science recognizes no other examples of ‘causal danglers,’ ontologically independent states with causes but no effects.”³²

Clayton’s Move from Emergentism to Dualist Theism

There are thus deep problems with Clayton’s emergentist account of mind. Even more serious difficulties occur as Clayton attempts to move from this *emergentist* account of mind to a *dualist* version of theism. Clayton attempts to move from emergence to transcendent dualism by arguing that there are things that even an emergentist account of the human person cannot explain.

The theistic emergentist...argues that the thought and action of *homo sapiens* (among other phenomena) confront us with certain predicates, qualities, and beliefs that are anomalous from the standpoint of natural law. Explaining these qualities and assessing the truth of these beliefs sets in motion an explanatory chain that eventually leads outside of natural science, and thus beyond the theoretical resources of emergence theory....The theistic account concludes to a conscious intentional being or force that preceded the evolutionary process and whose creative intentions led, however, indirectly, to the emergence of intelligent life.³³

³²David Papineau, *Thinking About Consciousness* (Oxford: Oxford University Press, 2002), 23.

³³Clayton, *Mind and Emergence: From Quantum to Consciousness*, 184.

What sorts of phenomena? In *Mind and Emergence*, Clayton highlights human rationality, the fact that human beings can form true beliefs about the world. “As reasoning agents we presuppose a fit between our beliefs and the external world. What must we postulate if we are to make sense of this core presupposition of human reason?”³⁴ Following a line of arguments developed by Alvin Plantinga, Clayton argues that theism is able to account for this aspect of human experience:

Plantinga’s broader philosophical project involves an additional claim: we do not finally have reason to trust the deliverances of our reason unless we postulate a self-conscious, rational creator who is benevolently disposed towards humanity, that is, one who intends for humans to form true beliefs and who creates them and the world such that, at least in most cases, this goal will be fulfilled. It is not enough that we postulate that we are epistemically ‘at home in the universe,’ as Nagel writes, we must actually *be* at home in the universe. And this will only be the case if the rational fit between mind and world was intentionally created.³⁵

Clayton is referring to Plantinga’s “evolutionary argument against naturalism.” The point of this argument is that evolutionary theory leads to skepticism regarding our ability to form true beliefs and thus undercuts the doctrine of naturalism. Because evolution is concerned with survival rather than truth, we cannot be in a position to know whether naturalism is true. This skeptical problem does not occur, however, when evolution is linked with theism. Plantinga, himself, presents the argument in the following manner:

Now according to traditional Christian (and Jewish and Muslim) thought, we human beings have been created in the image of God. This means, among other things, that God created us with the capacity for achieving *knowledge*—knowledge of our environment by way of perception, of other people by way of something like what Thomas Reid called *sympathy*, of the past by memory and testimony, of mathematics and logics by reason,

³⁴Ibid., 177.

³⁵ Ibid., 178.

of morality, of our own mental life, of God himself, and much more. And the above evolutionary account of our origins is compatible with the theistic view that God has created us in his image. So evolutionary theory taken by itself (without the patina of philosophical naturalism that often accompanies expositions of it) is not as such in tension with the idea that God has created us and our cognitive faculties in such a way that the later are reliable, that (as the medieval liked to say), there is an adequation of intellect to reality.

But if *naturalism* is true, there is no God, and hence no God (or anyone else) overseeing our development and orchestrating the course of our evolution. And this leads directly to the question whether it is at all likely that our cognitive faculties, given naturalism and given their evolutionary origin would have developed in such a way as to be reliable, to furnish us with most true beliefs.³⁶

Plantinga's argument offers a critique of naturalism on the basis of evolution and also proposes a theistic solution to the problem. Both the critique of naturalism and the theistic solution are not compelling. His argument fails to distinguish different kinds of cognition generated by evolution and how they may or may not be reliable. In particular, in order to undercut naturalism, Plantinga needs to show how *scientific* beliefs are unreliable. This would require a critique of the rudimentary forms of cognition that provide the platform for scientific beliefs as well as the actual methods and practices of science that refine and hone such beliefs. The naturalist may well be skeptical of the truth value of many kinds of human belief; she is committed only to the reliability of a certain set of scientific and philosophical beliefs. Further, in terms of the relationship between evolution and the formation of true beliefs, Plantinga overlooks the very real ways in which survival and truth can be linked. The two notions are not necessarily at loggerheads. After all, an organism that failed to represent its environment accurately

³⁶Alvin Plantinga, *Warrant and Proper Function* (Oxford: Oxford University Press, 1993), 229.

would not be expected to be around for long. In discussing the relationship of survival and truth in evolution, William Ramsey draws an analogy with camouflage:

Having a good camouflage is often a key adaptive feature that is favored by natural selection. However, since the quality of a given camouflage depends, in part, on the environment, a good camouflage does not supervene on an organism's intrinsic physical makeup. This is a fairly common case where a non-supervening property (having a good camouflage) enhances a creature's reproductive fitness and, hence, becomes a property for which there is considerable selection pressure. The irreducibility of a good camouflage in no way undermines its central importance to natural selection.

What all of this suggests is that there is a perfectly plausible way for truth to be a property that, although irreducible, is nevertheless favored by natural selection. Indeed, since some naturalists characterize beliefs as the "maps by which we steer," a slight modification of our earlier example shows us exactly how this would work. Suppose we have ten creatures competing for a scarce resource such as food, and suppose that only one of these creatures (Bob) possesses accurate beliefs concerning the whereabouts of the food. If all we want is an explanation of Bob's immediate motor behavior, then we need not appeal to the truth or falsehood of any of his beliefs. But if we want to know why Bob's behavior proves successful while his cohorts die out, then it clearly does matter that his behavior is generated by true beliefs....Hence, truth and reliability are exactly the sort of features for which there can indeed be considerable selection pressure.³⁷

Plantinga overlooks how truth can be adaptively significant and overlooks the difference between the claim that *some* of our beliefs may be unreliable and the claim that *all* our beliefs—including scientific beliefs—are unreliable.

In fact, there is indeed good evidence that some of our beliefs are unreliable, but this creates a problem for the theist and not the naturalist.³⁸ Claims about physical design

³⁷William Ramsey, "Naturalism Defended," in *Naturalism Defeated?: Essays on Plantinga's Evolutionary Argument against Naturalism*, ed. James K. Beilby (Ithaca: Cornell University Press, 2002), 18.

³⁸This evidence comes largely from scientific psychology and concerns everyday beliefs and decision making. Social psychologists like Richard Nisbett, Timothy Wilson, and Daniel Wegner have

provide an apt analogy. In such cases, claims about God as an intelligent designer are undercut by obvious design flaws in nature. Evolution is a mixed bag, and our evolved beliefs are a mixed bag as well. This mixed bag is exactly what one would expect from a naturalistic perspective. However, the theist argues that God has designed us to acquire true beliefs, surely the prevalence of false beliefs are an embarrassment to the theist not to the naturalist. There is also the issue of alternative religious epistemologies. The monotheistic solution that Plantinga offers is but one account; why assume that this account is the accurate account? Why should one appeal to Plantinga's Christian account of knowledge rather than say a Hindu account or Buddhist account?

There are significant issues with Plantinga's argument; however, even more problems occur in Clayton's appropriation of Plantinga's argument. In the evolutionary argument against naturalism, Plantinga is coy about his own beliefs about evolution and simply assumes evolution and theism are compatible. Plantinga, in fact, rejects evolutionary theory and also defends a robust version of substance dualism. In other places, Plantinga makes it clear that evolution and theism make strange bedfellows.³⁹ Because Clayton affirms both evolutionary theory and an emergentist account of mind,

probed the unreliability of our introspective beliefs in numerous empirical studies. For an overview see, Richard Nisbett, and Timothy Wilson "Telling More Than We Can Know: Verbal Reports on Mental Processes," *Psychological Review* 84 (1977); Daniel M. Wegner, *The Illusion of Conscious Will* (Cambridge, MA: MIT Press, 2002); Timothy D. Wilson, *Strangers to Ourselves: Discovering the Adaptive Unconscious* (Cambridge, MA: Belknap Press of Harvard University Press, 2002).

³⁹See Alvin Plantinga, "Advice to Christian Philosophers," *Faith and Philosophy: Journal of the Society of Christian Philosophers* 1 October (1984); Alvin Plantinga, "On Christian Scholarship," (1994). http://www.calvin.edu/academic/philosophy/virtual_library/articles/plantinga_alvin/on_christian_scholarship.pdf (accessed 7/31/09); Alvin Plantinga, "Materialism and Christian Belief," in *Persons: Human and Divine*, ed. Peter van Inwagen and Dean Zimmerman (Oxford: Oxford University Press, 2007).

he faces additional challenges that Plantinga is able to avoid. If evolution does have a tendency to produce unreliable beliefs, as Plantinga claims, how does God ensure the production of true beliefs throughout the evolutionary process? Indeed, why would God choose to create in this manner in the first place? Further, given Clayton's confidence in evolutionary-emergentist explanation, what reason do we have for thinking that such a framework could not account for our ability to form true beliefs?

In short, Plantinga's argument is problematic on its own, but it becomes even more problematic given Clayton's assumptions about evolution and emergence. Strong emergentism understands mind as "of a piece" with the natural world, and, this move can be understood as undercutting theism as Clayton himself acknowledges. Recall his quotation:

[T]he case for emergent mental causation is not by itself a case for the existence of God, divine action, an eternal soul, or life after death; it is not directly a theological conclusion at all. Indeed, in some ways it might seem to be an *anti*-theological conclusion, because it understands mental phenomena to be "of a piece" with physical phenomena, and because the supervenience relationship asserts a dependence of the mental life on its physical basis—indeed a high correlation between physical causes and mental effects—which is on the surface inconsistent with many parts of Christian teaching.⁴⁰

Given this, the bar is certainly high for a move to theistic dualism, but, for the reasons discussed above, Clayton's appropriation of Plantinga's evolutionary argument against naturalism is simply not compelling.

⁴⁰Philip Clayton, "Neuroscience, the Person, and God: An Emergentist Account," 206.

Experiential Religious Dualism: B. Alan Wallace

We have been addressing the implications of the scientific image of persons on religion. Discussion thus far has centered on the link between dualism and theism in the context of the Christian tradition. Christianity has figured particularly prominently in discussions of religion and science, to the point where discussions of “Christianity and science” are seen as synonymous with discussions of “religion and science.” Due to its radically different ontological vision and its typically pragmatic focus, the dialogue between Buddhism and science occurs on terms very different from those that define the Christianity and science dialogue.

Because Christianity has largely set the terms for religion and science discussions in the West, Buddhism has proved intriguing for some Western scientists and philosophers given its non theistic vision, radically different understanding of the self, methods for exploring consciousness, and ethical and transformative concerns. Philosopher of mind Owen Flanagan captures this general enthusiasm when he writes:

But it is worth pausing to reflect on the apparent accident that Buddhism, almost alone among the great ethical and metaphysical traditions, holds to a picture of persons that is uniquely suited to the way science says we ought to see ourselves and our place in the world.⁴¹

Indeed one might get the impression from such a quotation that Buddhism amounts to a kind of paleo-naturalism; however, as much as Christianity, Buddhism is fueled by a comprehensive soteriological and eschatological vision and this vision implies a certain ontology. God and soul do not figure in this ontological vision, but

⁴¹Owen Flanagan, *The Problem of the Soul: Two Versions of Mind and How to Reconcile Them* (New York: Basic Books, 2002), 210.

immaterial consciousness does. As the fourteenth Dalai Lama has written recently in his book promoting “the convergence of science and spirituality”:

There is no reason to believe that the innate mind, the very essential luminous nature of awareness has neural correlates, because it is not physical, not contingent upon the brain...I feel that on a more subtle level of consciousness, brain and mind are two separate entities.⁴²

A Buddhist programme for religion and science has been outlined in recent works by one of the Dalai Lama’s translators, the American Buddhologist, B. Alan Wallace.

While Christian theological discussions of religion and science tend to take the doctrinal dimension as their starting point, Wallace’s approach centers around the experiential and ethical dimensions of religion. But the experiential and ethical dimensions of religion are not separate from metaphysical questions about what exists in Wallace’s thought; rather, it is precisely through experiential-ethical practices that one discovers the deepest *truths* about the nature of conscious mind. In particular, a notion of immaterial consciousness is central in Wallace’s understanding of Buddhism, and Wallace’s Buddhist approach bears interesting affinities with classical theistic dualism in its critique of naturalism and its conviction that the origin of the psyche cannot be found in the brain.

In classical theism, immaterial consciousness is discussed in reference to God and the soul. Wallace’s thought, true to its Buddhist roots, focuses neither on God nor the soul, but does argue that a notion of immaterial consciousness is central to secure key Buddhist beliefs:

While all Buddhist schools refute the existence of an immutable, unitary, independent self, this does not mean that there is no continuity of an

⁴²Dalai Lama and Daniel Goleman, “On the Luminosity of Being,” *New Scientist* 178, no. 2396 (2003): 42.

individual stream of consciousness after death. Indeed, this seems to be indispensable if Buddhists are to accept the theory of karma and reincarnation taught in the Pali Canon, which also refers to the existence of an intermediate phase of conscious experience following death and prior to one's next incarnation.⁴³

Wallace also argues that this immaterial form of consciousness—known as substrate consciousness—is discoverable through meditative experience:

[T]he hypothesis of the substrate consciousness rejects both Cartesian dualism...and the belief that the universe is exclusively physical. Moreover, it may be put to the test of experience, regardless of one's ideological commitments and theoretical assumptions.⁴⁴

As the above quotation illustrates, there are very significant methodological differences in the arguments of the classical theists and in the arguments of Wallace. Classical theists tend to focus on the doctrinal aspects of religion and attempt to secure Christian doctrinal claims by means of theological and philosophical arguments. Wallace's approach places much more emphasis on the role of experience. Wallace calls his approach, "contemplative science," and he views it as a kind of via media between doctrinal religion and materialist science.

The human mind cannot be thoroughly comprehended only through the scientific examination of the brain and behavior; the human soul cannot be fathomed only on the basis of divine revelation. Science is not equipped to explore the spiritual dimensions of existence, for its tools have been designed to measure physical processes. Science and religion may yet prove to be complementary, but only if adherents of both return to the primacy of experience.⁴⁵

⁴³B. Alan Wallace, *Mind in the Balance: Meditation in Science, Buddhism, and Christianity* (New York: Columbia University Press, 2009), 97.

⁴⁴Ibid., 92.

⁴⁵B. Alan Wallace, *Contemplative Science: Where Buddhism and Neuroscience Converge* (New York: Columbia University Press, 2007), 44.

A central claim in Wallace's thought is that the West has not developed an adequate science of consciousness largely because it views consciousness through the cramped optics of scientific materialism:

[T]he West presently has no *pure* science of consciousness that reveals the nature, origins, and potentials of this natural phenomenon, and it similarly lacks an *applied* science of consciousness that reveals means for refining and enhancing consciousness and thereby achieving *eudaimonia*. But this does not necessarily imply that all other human civilizations throughout history have been equally deficient.⁴⁶

The issue of the origin of the psyche is a central focus in Wallace's contemplative science project, especially the claim that the mind cannot be identified with the brain. Wallace offers both positive and negative arguments concerning the origin of the psyche. This section examines and critiques these arguments.

Wallace's Negative Arguments about Naturalism and a Naturalistic Explanation of Consciousness

Much of Wallace's work is premised on the notion that scientific naturalism or materialism cannot adequately account for the emergence and existence of the conscious mind. Scientific materialism is in fact a "dogma" that obscures and prevents the development of an authentic science of consciousness.⁴⁷ Wallace understands scientific materialism in light of five central notions: objectivism, reductionism, the closure principle, monism, and physicalism.⁴⁸ The principle of objectivism requires that science

⁴⁶Ibid., 58.

⁴⁷See B. Alan Wallace, *The Taboo of Subjectivity: Toward a New Science of Consciousness* (Oxford: Oxford University Press, 2000), 145-175.

⁴⁸Ibid., 17-27.

deal with empirical facts that are testable by empirical methods and verifiable by third-person means. The closure principle holds that the physical world is causally closed: there are no causes other than physical causes. Monism, reductionism, and physicalism are closely related: reality is composed of fundamental stuff—i.e., the stuff defined by physics and macro-phenomena can be understood in light of the causal work being done at lower levels.

However successful these principles might have been in launching modern science, Wallace argues that they have been disastrous for the study of mind because they entirely omit a first-person subjective perspective and lead people to believe that the brain is the exclusive origin of the conscious mind.

Are the methods devised for the study of objective, physical phenomena sufficient for the scientific study of subjective, mental phenomena? At this point, all answers are expressions of faith, for it is obvious that physical and biological sciences have not yet comprehensively explained the origins, nature, or causal efficacy of consciousness or any other mental phenomenon. Great advances have been made recently in discovering the neural correlates of an increasing range of mental processes, but none of these explains the “hard problem” of how these physical events give rise to subjective experience. But there are compelling reasons for skepticism about the neurobiological reduction of the mind to the brain.⁴⁹

Wallace’s negative arguments mirror anti-naturalist arguments that have been discussed throughout the dissertation. The “compelling reasons for skepticism” amount to an assertion of the Correlation Claim regarding the relationship between mind and brain. In light of the arguments developed thus far in this dissertation, such skepticism is misplaced. We should be skeptical of principled arguments against the bio-psychological explanation of consciousness, not of the nascent science of the mind-brain. Wallace

⁴⁹Wallace, *Contemplative Science: Where Buddhism and Neuroscience Converge*, 15.

repeatedly claims that there is no explanation of the hard problem of consciousness, and this is true to a point, but Wallace must establish that there cannot *in principle* be a naturalistic solution to the hard problem, and surely this is an imprudent move given the relative youth of both neuroscience and consciousness studies and given the resources naturalism has for explaining subjectivity.

Brain-consciousness science is not a completed science. Neuroscience itself is an incredibly young discipline. The neuron doctrine was established a mere century ago. Serious scientific research on consciousness is less than two decades old and still in a pre-paradigmatic phase. Given these factors it is hardly surprising that there is no solution to the hard problem currently available. Is Wallace really ready to assert that we cannot have a bio-psychological explanation of consciousness at this early stage of consciousness research? There is a massive amount of evidence from the history of science that is a very unwise choice. Chapter One spent an extensive amount of time with the most philosophical rigorous claims that consciousness could never be explained in terms of brain processes: they simply are not compelling. They assume that our concepts of both consciousness and the brain are fixed; however, given the youth of neuroscience and the infancy of scientific work on consciousness, this is not a particularly prudent judgment.

Wallace also presents a rather monolithic analysis of naturalism and underestimates the degree to which naturalism can accommodate subjectivity. Wallace's thinking is mired in the same kind of dichotomous thinking that has been critiqued throughout this dissertation. Wallace seems to equate naturalism or materialism with the

view that only objective physical facts exist; he opposes this with the dualist view that there are objective physical facts and subjective nonphysical facts. However, explaining away consciousness or placing it outside the physical world are not the only options. A robustly naturalistic perspective can accommodate both objective physical facts and subjective physical facts. Subjective physical facts are objective physical facts experienced “from the inside” with the characteristics of transparency and perspectivalness. Wallace does not discuss the range of positions open to the naturalist nor does he consider in detail any scientific approach to consciousness or the considerable scientific problems dualism generates.

In the position of this dissertation, the brain-consciousness problem is epistemic and not metaphysical. Conscious mental states can be understood in terms of objective physical facts and subjective physical facts. Subjective physical facts are not facts that hover somehow “over and above” objective physical facts. They are epistemologically novel, but not ontologically novel. They are objective physical facts viewed from a transparent, first person perspective. Consciousness and the conscious self are virtual realities created by the brain. Such a position forms the hard core of a scientific research program and takes subjectivity seriously while cohering better with established scientific knowledge about the world. Such a position may in fact turn out to be wrong empirically, but it is surely too early to determine that, and principled arguments that resist the claim that the mind can be explained in terms of the brain simply are not compelling.

Wallace's Case for Buddhist Religious Dualism

Wallace's arguments about consciousness go far beyond negative property dualist arguments about the irreducibility of conscious mental properties. He makes very robust metaphysical claims about the nature of substrate consciousness. Wallace's skepticism about the scientific study of consciousness is matched by a confidence about Buddhist claims regarding the substrate consciousness. Wallace argues that fundamental truths about the nature and origin of consciousness can be uncovered through meditative practice. Buddhist thought understands mental activity (Pali: *javana*; Sanskrit: *chitta*) as grounded in a more fundamental ground state of consciousness known as the ground of becoming (*bhavanga*) in early Theravada Buddhism and as substrate consciousness (*alayavijnana*), in the Great Perfection (*Dzogchen*) school of Tibetan Buddhism. This is the individual stream of consciousness that carries on from one lifetime to the next. Buddhists describe its fundamental characteristics as luminosity and cognizance. Thus in human being's consciousness is *related to* brains and bodies, but it emerges from the substrate consciousness. According to the Great Perfection tradition of Buddhism, this ground state of consciousness is normally inaccessible, as it is manifested mainly in deep sleep. However, with the cultivation of Buddhist meditative practice or *samadhi*, the substrate consciousness can be actively experienced.

Buddhist contemplatives claim that with the achievement of a highly advanced degree of *samadhi* known as *samatha* or meditative quiescence, one gains experiential access to the relative ground state of consciousness known in the Great Perfection (Dzogchen) school of Tibetan Buddhism as the "substrate consciousness (*alayavijnana*). This, they claim, is the individual stream of consciousness from which the psyche and all the physical senses emerge. According to their findings, the

psyche is *conditioned* by the body and its physical interaction with the environment, but it *emerges* from the substrate consciousness.⁵⁰

This state of meditative quiescence is not simply a powerful experience, it discloses the fundamental nature of consciousness, most importantly, the notion that “consciousness can only come from consciousness”:

Contemplatives who have explored this immaterial dimension of reality have discovered a principle of conservation of consciousness that manifests in every moment of experience. No constituents of the body—in the brain or elsewhere—transform into mental states and processes. Such subjective experiences do not emerge from the body, but neither do they emerge from nothing. Rather, all objective mental appearances arise from the substrate, and all subjective mental states and processes arise from the substrate consciousness.⁵¹

While substrate consciousness is clearly a notion of immaterial mind, Wallace is careful to distinguish it from the notion of soul:

While this description of the substrate consciousness may appear to be a Buddhist version of an immaterial soul, it is important to note the differences between this experientially based account and various philosophical and theological speculations about the soul. Contemplatives who have achieved *samatha* commonly depict this dimension of consciousness as a stream of arising and passing moments of awareness, so it is not a single entity persisting through time, nor is it unchanging. Moreover, it influences the psyche and is conditioned by physical and mental events, so it is not independent.⁵²

In turn, the individual stream that is substrate consciousness is grounded in yet a third dimension of consciousness, primordial consciousness (*jnana*) or the Buddha-nature. As

⁵⁰Wallace, *Contemplative Science: Where Buddhism and Neuroscience Converge*, 17.

⁵¹Wallace, *Mind in the Balance: Meditation in Science, Buddhism, and Christianity*, 95-96.

⁵²Wallace, *Contemplative Science: Where Buddhism and Neuroscience Converge*, 17-18.

the psyche emerges from substrate consciousness, so all streams of substrate consciousness originate from primordial consciousness.

This is regarded in the Mahayana Buddhist tradition as the ultimate ground state of consciousness....The realization of this state of consciousness is said to yield a state of well-being that represents the culmination of the Buddhist pursuit of eudaimonic well-being, knowledge, and virtue. With such insight, it is said that one comes to understand not only the nature of consciousness but its relation to reality as a whole....The substrate consciousness can allegedly be ascertained with the achievement of the advanced states of *Samadhi*, whereas primordial consciousness can be realized only through the cultivation of contemplative insight (*vipashyana*). Thus, Buddhism postulates this dimension of awareness not as a mystical theology, but as a hypothesis that can be put to the test of immediate experience through advanced contemplative training open to anyone, without a leap of faith that violates reason.⁵³

Thus meditative practice is seen as providing not only experiential-ethical transformation but insight into the nature of consciousness itself. Wallace claims that this experience is open to all who engage in meditative practice, thus it can be put to the test of experience. He also points to cross-cultural and interreligious convergence regarding contemplative-mystical experience:

There are many important differences between Buddhist and Christian theories of consciousness, and between scientific and contemplative theories of space. But in the midst of these doctrinal and theoretical differences, there may be a hidden ground on which these diverse traditions ultimately converge. If so, I believe they are converging on the most important truth that can be known and experienced.⁵⁴

Wallace holds that different religious traditions share a mystical-contemplative-experiential core, and that this experiential core yields insight into “the most important

⁵³Ibid., 50.

⁵⁴Ibid., 24.

truth that can be known and experienced.” A great deal of Wallace’s thought sounds like a version of perennial philosophy that sees all religions united around a common mystical-contemplative core. Wallace himself makes this association at one point in his writings:

This integrative view, asserting a common ground to some of the deepest contemplative insights of diverse traditions despite the differences in their institutional doctrines, has come to be known in modern times as the perennial philosophy. Under the domination of so much postmodern thinking in the current academic study of religion, this is presently not in vogue, but it has been endorsed to varying degrees by many influential religious scholars, including William James, Rudolf Otto, Aldous Huxley, Ninian Smart, Huston Smith, and Robert Forman.

Over the centuries, Buddhism, like Christianity, has produced a rich diversity of ways of viewing ultimate reality, the phenomenal world, and human nature. But there may be a luminous common ground in their deepest contemplative insights that is temporarily veiled by the biases and obscurations of the human mind.⁵⁵

An important question for perennial philosophy is what “the most important truth” or “luminous common ground” consists in. Perennialists typically hold that different religious traditions share a core common experience that transcends the doctrinal particularities of individual religious traditions. Wallace here makes somewhat of a different claim. He affirms a “luminous common ground” shared between diverse religious traditions, but links this common ground specifically to Buddhist claims regarding the substrate consciousness. In other words, he offers an affirmation of universal mystical-contemplative experience, but links this experience with highly particular Buddhist beliefs. Wallace claims this experience of the substrate consciousness includes the experiential verification of Buddhist beliefs regarding karmic rebirth. “Belief

⁵⁵Ibid., 107-108.

in reincarnation is prevalent in all schools of Buddhism, initially stemming from the Buddha's experience of enlightenment....This contemplative discovery of the existence of past lives has allegedly been replicated by many generations of Buddhist meditators who have developed samadhi and used it to explore the nature of origins of consciousness."⁵⁶ Wallace also argues that claims regarding reincarnation are amendable to scientific research:

While Christian and materialist beliefs regarding what happens at death remain largely unquestioned within their respective communities, theories of reincarnation do lend themselves to experiential investigation and rational analysis. Over the past forty years, scientists have identified and studied several thousand cases of young children from all over the world who have accurately reported alleged memories of their past lives. The late Ian Stevenson, professor emeritus of psychiatry and the former director of Personality Studies at the University of Virginia, pioneered this line of research and wrote extensively on it....More scientifically compelling evidence comes from the field of "near death" and "out-of-body-experiences."⁵⁷

Wallace's contemplative science project contains a number of claims about meditation and meditative experience. Indeed claims about meditative practice and experience are at the crux of the Buddhism-mind science dialogue. A firestorm erupted over the invitation of the Dalai Lama to speak at the 2005 meeting of the Society of Neuroscience. Was the Dalai Lama discussing a technique with certain empirically assessable physical and mental health benefits, or was he rather offering a tacit form of Buddhist apologetics? These same sorts of ambiguities occur throughout the writings of Wallace, and an analysis of the Buddhism-mind science dialogue involves sorting out and

⁵⁶Wallace, *Mind in the Balance: Meditation in Science, Buddhism, and Christianity*, 105.

⁵⁷Wallace, *Contemplative Science: Where Buddhism and Neuroscience Converge*, 106-107, 110.

analyzing various claims regarding Buddhist meditation and experience. The complexities involved in making sense of claims regarding meditation in the Buddhism-mind science are described well by Donald Lopez:

The claim here is that Buddhist meditation works. However, in order to understand the laboratory findings, such a claim requires that one first identify what is *Buddhist* about this meditation, describe what the term *meditation* encompasses in this case, and explain what *works* means, especially in the context of the exalted goals that have traditionally been ascribed to Buddhist practice. Although these goals are numerous and variously articulated across the tradition, it can be said that their ultimate aim is not self-help but a radical reorientation toward the world—and in many articulations, a liberation from it—either for oneself or for all beings.⁵⁸

Wallace's writings contain a number of claims regarding meditation. Proper analysis of Wallace's work involves differentiating the numerous claims he makes about meditation and meditative experience. There are at least seven different claims regarding Buddhist meditation within Wallace's writing: (1) Meditative techniques have physical and mental health benefits; (2) Meditative techniques have moral benefits; (3) Meditative techniques can be used by all regardless of their cultural and religious background; (4) Contemplative/meditative experiences have interreligious and cross-cultural similarities; (5) Meditation shows that consciousness is nonphysical; (6) Meditative practice and experience is the best way to bridge the concepts of religion and science; (7) Meditation shows that the claims of Tibetan Buddhism concerning the substrate consciousness are true.

⁵⁸Donald Lopez, *Buddhism and Science: A Guide for the Perplexed* (Chicago: University of Chicago Press, 2008), 207.

There is nothing, *in principle* that a naturalist would object to in the first four claims, significant objections can be raised about the last three claims, however.

One very important claim that Wallace makes with regard to philosophy of mind is the claim that that meditative experience reveals consciousness to be immaterial.

Certainly, the prior arguments of this dissertation would caution against a premature move from the *introspective experience* of consciousness to metaphysical claims about the *nonphysical nature* of consciousness. Conscious experience may certainly *seem* to be nonphysical, but that doesn't mean that it is nonphysical, and there are good scientific and philosophical reasons for thinking that consciousness is physical. Dualism does not cohere well with the findings of physics, evolutionary biology, and neuroscience, and positive arguments for dualism fail.

Moreover, as Chapter Two argued, the sense that consciousness does not seem to be physical can be accounted for in completely naturalistic terms. There are a number of neurological and psychological factors that bias human beings toward dualistic thinking. Neurologically, we do not see brain activity when we are thinking, nor do we sense the brain. Further, the brain does not appear in our conscious thought; all of our experience is “out-of-brain” experience, but that does not mean that it is literally out of the brain. If the claims of Antti Revonsuo, Thomas Metzinger, and Olaf Blanke are correct, the virtual experience of being “outside our brains” and even “outside our bodies” is entirely amenable to scientific study. Earlier sections in Chapter Two explored how psychological factors also contribute to a dualist bias. We have two separate psychological systems for navigating the physical world and social world. One by-product of these evolutionary

adaptations is that they predispose us to think of the world in terms of body and soul. Death also places impossible constraints on our theory of mind: we cannot know what it is “like” to be dead, i.e., consciously simulate a permanent state of unconsciousness. This coupled with the fact that some mental states are easier to imagine the absence of than others, dispose us to strongly dualist afterlife beliefs and to represent dead agents’ minds in certain ways. The notion of minds enjoying an intermediate state between rebirths would seem to fit nicely into this framework.

Contemporary scientific research tells us that we have a predisposition to dualist thought. A separate strand of psychological research pioneered by Richard Nesbitt, Timothy Wilson, and Daniel Wegner points to the fallibility of a number of our common sense introspective judgments. Given this research, plus the enormous philosophical and scientific problems associated with dualism, claims that consciousness introspection resolve the mind-body problem in favor of dualism should be greeted with a certain amount of suspicion. Such research cautions against a slide from folk phenomenology and folk psychology to metaphysical claims about dualism. If this is true of ordinary states of consciousness, it certainly should be true of altered states of consciousness, however profound.

Wallace also makes a number of thick claims regarding contemplative-meditative experience. In his writing on contemplative and meditative practices, Wallace links four important claims: (1) that the experiential dimension and the mystical-contemplative strands of the world’s religious traditions represent these traditions at their most authentic; (2) that mystical-contemplative experiences provide a common ground

between different religious traditions; (3) that these experiences are metaphysically significant in that they are disclosive of a transcendent immaterial reality; and (4) that this reality is in fact the substrate consciousness described in Tibetan Buddhism.

One very common conception of religion in the West today locates the essence of religion in an experiential core and sees this experiential core as something shared across diverse religious traditions. This view is frequently referred to as the Perennial Philosophy, following Aldous Huxley's book by that name. The claims of Perennial Philosophy can be understood both descriptively and normatively. The descriptive claim is that (some) religions locate the essence of religion in experience. The normative claim is the claim that an experiential focus provides the most adequate approach to the study of religion (or religion and science). There are reasons to be skeptical of both the descriptive and normative claims of Perennial Philosophy.

History provides one reason for skepticism regarding the descriptive claims of Perennial Philosophy. A cursory glance at the history of Christianity provides an example. While there is a very significant strand of neo-platonic mysticism in Christianity, the tradition also contains a long-standing suspicion of religious and mystical experience that stretches from Paul to Martin Luther to the twentieth century Protestant theologian Karl Barth and his followers. Even where the mystical-experiential element of religion is affirmed (e.g., in Eastern Orthodoxy and Roman Catholicism) it is typically regarded as a divine gift limited to a select few. Further, such traditions have constantly stressed the content (i.e., the Christocentric nature) of even the most apophatic forms of Christian mysticism.

Recent historical analyses of religious experience by Wayne Proudfoot, Robert Sharf and others argue that it is in the context of modernity that claims about religious experience become elevated.⁵⁹ Experience rises to prominence as a result of modern scientific, historical, and cultural challenges to theology and the study of religion more generally. By locating the essence of religion in experience—as opposed to scripture or doctrine—theologians could evade the corrosive effects of modern historical and scientific inquiry:

By emphasizing the experiential dimension of religion—a dimension inaccessible to strictly objective modes of inquiry—the theologian could forestall scientific critique. Religious truth claims were not understood as pertaining to the objective or material world, which was the proper domain of science, but to the inner spiritual world, for which the scientific method was deemed inappropriate...⁶⁰

The notion that all religions shared a common experiential core also provided an apt solution to an increasing awareness of cultural pluralism:

By the twentieth century it had become difficult for Christian theologians to simply ignore the existence of non-Christian traditions, much less to smugly assert Christian superiority. But to take other traditions seriously entailed the risk of rendering Christianity merely one of several competing systems of belief. In privileging religious experience, theologians could argue that all religious traditions emerged from, and were attempts to give expression to, an apprehension of the divine or the ultimate. Differences in doctrine and forms of worship are to be expected due to vast differences in linguistic, social, and cultural conditions. What is key, however, is that as a response to a fundamentally human (and thus pancultural and ahistorical) sense of the transcendent, all religions could lay *some* claim to

⁵⁹Wayne Proudfoot, *Religious Experience* (Berkeley: University of California Press, 1985); Robert Sharf, "Buddhist Modernism and the Rhetoric of Meditative Experience," *Numen* 42 (1995); Robert Sharf, "Experience," in *Critical Terms for Religious Studies*, ed. Mark C. Taylor (Chicago: University of Chicago Press, 1998).

⁶⁰Sharf, "Experience," 95.

truth. This allowed Christian theologians to affirm the validity of Christian revelation without necessarily impugning their non-Christian rivals.⁶¹

The historical, social, and natural sciences presented a challenge not only to Christian theology, but to the academic study of religion more generally. If religion was something that could be studied historically, socially, psychologically, and biologically, why was there a need for a separate discipline of “religious studies”? Many scholars of religion justified the existence of their discipline precisely through appeal to the *sui generis* character of religious experience. As Sharf notes, “By construing religion as pertaining to a distinct mode of ‘experience,’ the scholar of religion could argue that it ultimately eludes the grasp of more empirically oriented disciplines.”⁶²

This genealogy of the notion of religious experience does not rule out that claim that experience may be a fruitful approach to the study of religion in the present, but it does caution against an anachronistic reading of premodern religious traditions. Here one might raise the objection that while the “turn to experience” may be relatively recent in the West, meditative-contemplative experience looms much larger in Eastern traditions like Hinduism and Buddhism. Indeed, these are precisely the claims that Wallace seems to be making with regard to Buddhism. In light of this, it is certainly significant that a number of Buddhologists caution against such a reading of premodern Eastern traditions. Referring specifically to the issue of meditation, Donald Lopez notes:

It is useful to recall that the vast majority of Buddhists over the course of Asian history have not practiced meditation. It has traditionally been

⁶¹Ibid., 96.

⁶²Ibid., 95.

regarded as something that monks do, indeed that only some monks do; the monastic codes make repeated reference to the needs of meditating monks, suggesting that they represented a group of specialists within the monastic order....In the Theravada cultures of Sri Lanka and Southeast Asia, there has been a long tradition of dividing monastic practice into two categories: the vocation of texts and the vocation of meditation. In commentaries dating from as early as the fifth century, a preference was expressed for the former....And there are major forms of Buddhism, most notably the Pure Land traditions, in which the practice of meditation does not play a central role at any state of the path...⁶³

In a series of provocative articles, Robert Sharf has argued that this focus on experience in the modern West greatly influenced the reading of Eastern religious traditions as grounded in experience from their inception. He cautions against reading premodern Buddhist texts as exalting personal experience.

The notion that Asian religions are more experientially rooted than their Western counterparts is one of those truisms so widely and unquestioningly held that corroboration of any kind is deemed superfluous. But when we turn to premodern Asian sources, the evidence is ambiguous at best. Take, for example, the many important Buddhist exegetical works that delineate the Buddhist *marga* or “path to liberation”...These texts are frequently construed as descriptive accounts of meditative states based on the personal experiences of accomplished adepts. Yet rarely if ever do the authors of these compendiums claim to base their expositions on their own experience. On the contrary, the authority of exegetes such as Kamalasila, Buddhaghosa, and Chih-i lay not in their access to exalted spiritual states but in their mastery of, and rigorous adherence to, sacred scripture.⁶⁴

Sharf also cautions against too close a link between meditation and personal experience:

The notion that meditation is central to Asian religious praxis might seem to support the thesis that Asian traditions exalt personal experience. But here too we must be cautious: contemporary accounts of Asian meditation typically *presume* that they are oriented toward meditative experience, and thus such accounts must be used with considerable caution. Besides, while

⁶³Lopez, 208.

⁶⁴Sharf, "Experience," 99.

meditation may have been esteemed in theory, it did not occupy the dominant role in monastic and ascetic life that is sometimes supposed. (This point is often overlooked by scholars who fail to distinguish between prescriptive and descriptive accounts.) Even when practiced, it is by no means obvious that traditional forms of meditation were oriented toward the attainment of “extraordinary states of consciousness.” Meditation was first and foremost a means of eliminating defilement, accumulating merit and supernatural power, invoking apotropaic deities, and so forth. This is not to deny that religious practitioners had *experiences* in the course of their training, just that such experiences were not considered the goal of practice, were not deemed doctrinally authoritative, and did not serve as the reference points for their understanding of the path.⁶⁵

The elevation of personal experience in fact happens through Eastern scholars thoroughly immersed in Western thought:

The valorization of experience in Asian thought can be traced to a handful of twentieth-century Asian religious leaders and apologists, all of whom were in sustained dialogue with their intellectual counterparts in the West. For example, the notion that personal experience constitutes the heart of the Hindu tradition originated with the prolific philosopher and statesman Sarvepalli Radhakrishnan. Like his European and American predecessors, Radhakrishnan argued that “if philosophy of religion is to become scientific, it must become empirical and found itself on religious experience” and “it is not true religion unless it eases to be a traditional view and becomes personal experience.”⁶⁶

There are philosophical as well as historical issues with the modern turn to religious experience. The modern turn to experience typically involves claims that religious traditions share experiential “luminous common ground” despite their significant doctrinal and cultural differences. In the twentieth century, this view came to be known as the perennial philosophy. The rendering of experience in perennial philosophy is deeply problematic and has come under much fire as the result of

⁶⁵Ibid.

⁶⁶Ibid.

developments in twentieth century thought. As Sharf and other have noted, the rendering of experience in perennial philosophy rests upon patently Cartesian assumptions:

The notion that the referent of the term “experience” is self-evident betrays a set of specifically Cartesian assumptions, according to which experience is held to be immediately present to consciousness....The rhetoric of religious experience, predicated as it is on Cartesian dualism, allowed scholars to distinguish the universal experiential ground of religion on the one hand, and its diverse culturally bound manifestations on the other, created an opposition that recapitulates the classical Cartesian bifurcation of mind and matter.⁶⁷

Twentieth century thought called attention to the cultural-linguistic mediation of thought and experience, and scholars advocating constructivist positions in both religious studies and theology have offered extensive critique of perennial philosophy and the notion of universal unmediated experience.⁶⁸ Constructivists provide a trenchant critique of the perennial philosophy and its Cartesian assumptions, but constructivism itself risks falling into a sort of dualism in construing human beings as socially constructed blank slates operating independently of brain and body. The problem with constructivism is not its stress on cultural-linguistic mediation; it is that it fails to link this mediation with the brain and body. Constructivist positions go wrong not in taking mediation seriously, but rather by not taking mediation seriously enough. This is not the place to develop a constructive account of religious experiences, but surely such an account needs to attend to what might call the “double mediation” of brain and body, on the one hand, and culture and language on the other. Such a double mediation involves a complex and reciprocal relationship between brain and body and culture and language. What is significant is that

⁶⁷Sharf, "Buddhist Modernism and the Rhetoric of Meditative Experience," 229.

⁶⁸See especially Steven Katz, ed. *Mysticism and Philosophical Analysis* (New York: Oxford University Press, 1978); Steven Katz, ed. *Mysticism and Religious Traditions* (New York: Oxford University Press, 1983).

such an account places one in a position to posit potential cross cultural similarities in religious experience without resorting to the metaphysics of the perennial philosophy. Commonalities in experiences deemed religious may be due to the much more mundane fact that human beings share a common biology. Whatever the case, contemporary historical, philosophical, and scientific analysis of the notion of religious experience gives one ample cause for skepticism regarding Wallace's claim that religion and science are best bridged by a return to the primacy of experience. An adequate account of religious experience surely needs to attend more to the cultural-linguistic and biological mediation of religious experience.

Wallace's experiential focus is closely linked with a commendable concern to address issues of religious diversity, and Wallace frequently points to experiential "common ground" between Buddhism and Christianity in particular. However, there is a deep ambiguity in Wallace's treatment of experience and religious diversity. Wallace affirms a universal experiential common ground between religious traditions but links that experiential common ground to very particular Buddhist claims regarding the substrate consciousness and karmic rebirth. Many who advocate forms of the perennial philosophy do so as a response to the challenge of religious diversity. Such perennialists also tend to be pluralists with regard to religious epistemology, claiming that no one particular religious tradition has a privileged explanatory purchase on religious truth or experience. It is important to see the difference between this pluralist stance and Wallace's position. Wallace's stance is not pluralism but a version of what has been called inclusivism. The inclusivist argues that there may well be a common experience

that links religious traditions, but that one tradition enjoys a privileged explanatory purchase with regard to those experiences. Wallace seems to assign Buddhism such a privileged rule and does not seem to regard Christianity and Buddhism as two equally valid paths, as other perennialists are apt to do. What is difficult to see is why one religious tradition should enjoy such an explanatory privilege.

It is hardly surprising that Buddhist cultures would generate powerful experiences of the substrate consciousness, including experiences involving the notions of karma and rebirth. What is much more dubious is the claim that Christian, Jewish, and Muslim mystics are also experiencing the substrate consciousness in spite of what they themselves claim to be experiencing. Wallace's linking of universal religious experience with the Buddhist notion of the substrate consciousness raises perplexing issues with regard to religious diversity and interreligious dialogue. The claim that one religious tradition is positioned to explain more adequately the religious experiences of another tradition strikes many today as both imperialistic and arrogant, as do claims that one religious tradition has a unique grasp of the truth. An awareness of cultural diversity contributes to this impression, but, the rise of modern science also contributes to skepticism regarding religious claims. Science has not only problematized religious claims about the world; it has also been able to develop a method for adjudicating differences and transcending cultural particularities in a way that religion simply has not been able to do. There are good reasons for accepting evolution over intelligent design theory, and for accepting the heliocentric versus the geocentric theory. While the theory of evolution and the heliocentric theory were developed in highly particular cultural

contexts they involve empirical claims about the world and are articulated in a manner that is publically accessible and verifiable. Religious claims to truth typically lack this empirical and public status.

Religious believers themselves recognize the epistemic status of science, and frequently appeal to science as part of an apologetic defense of their particular religious tradition. Wallace, for example, seems to suggest that science can potentially provide a verification of Buddhist beliefs. This specifically occurs with reference to belief in reincarnation/rebirth. It is also not surprising the Hindu-Buddhist cultures saturated in beliefs about reincarnation/rebirth would produce individuals claiming to have experienced or lived an earlier life. What is dubious is the claim that the truth of reincarnation is the best explanation of these experiences. Wallace points to the work of University of Virginia psychiatrist, Ian Stevenson, to add empirical support to his reincarnation claims. Wallace simply asserts that the scientific community fails to take this work seriously, and fails to engage any of the massive critical commentary that has accumulated around Stevenson's work.⁶⁹ The same is true for Wallace's appeal to out-of-body experiences (OBE) and near-death experience (NDE). These are interesting altered states of consciousness, and they are fully amenable to empirical study and research. None of this research however indicates that such experiences are direct evidence of the existence of an immaterial soul or an immaterial substrate consciousness. In fact, the ease at which bodily self-representation can be manipulated neurologically and

⁶⁹For an overview of such critiques see Paul Edwards, *Reincarnation: A Critical Examination* (New York: Prometheus Books, 1996).

psychologically, would seem to support a naturalistic, bio-psychological explanation of such phenomena.

The purpose of this chapter was to highlight the implications of the demise of mind-body dualism for religious thought. Sophisticated attempts to retain forms of religious dualism, such as those of Philip Clayton and B. Alan Wallace, are simply not compelling. The implications of the demise of dualism for dialogue between religion and science is discussed in the next chapter.

CHAPTER FOUR

SHIFTING THE RELIGION AND SCIENCE CONVERSATION: BIO-PSYCHOLOGICAL EXPLANATION OF MORALITY AND ITS IMPACT UPON METAETHICS AND RELIGION

Recap of the Dissertation Argument

There is currently a wild diversity of positions on the relationship between religion and science. This dissertation seeks to make some sense of that wild diversity by looking at a particular kind of science: the science of the mind-brain. One of its guiding insights is that the mind sciences surface in a particularly acute way one of the most foundational issues at stake in discussions of the relationship between religion and science: the issue of the nature of mind and the debate between dualism and naturalism. Resolving the debate between dualism and naturalism establishes important constraints for thinking about the relationship between religion and science and for a constructive approach to religion more generally.

Chapters One, Two, and Three focused on various aspects of the debate between dualism and naturalism. In Chapter One, I made the case that dualism is a fundamental aspect of religious traditions like Christianity and Buddhism and continues to be present in much contemporary religious and theological literature that seeks a harmonious integration of religion and science. Many religious traditions understand certain aspects of reality as literally spiritual, i.e., not part of the physical world described by the natural

sciences, and many academic theologians interested in religion and science are prepared to defend these dualist beliefs as metaphysical claims about the nature of the human person and the nature of the cosmos. I called this stance “religious dualism.” Religious dualism includes forms of classical theistic supernaturalism, but it extends far beyond this. Many theologians engaged in the religion and science dialogue repudiate supernaturalist theism while retaining a theistic dualism. Religious dualists all hold versions of the thesis that consciousness can exist without neural correlates or any physical substrate at all.

Having established the foundational role of dualism within much religious thought, I turned to the evaluation of these dualist claims on the basis of the contemporary philosophy and science of the mind-brain. My concern in Chapters One and Two was to assess the notion of dualism as a doctrine about human persons. In those chapters, I attempted to assess what light the contemporary mind sciences could shed on the traditional mind-body or consciousness-brain problem. Chapter One presented the debate between naturalism and dualism as a debate about the adequacy of a bio-psychological explanation of consciousness and argued that principled dualist, mysterian, functionalist/autonomy of psychology objections to a vertically integrated bio-psychological approach to consciousness are simply not compelling. Such claims are typically rooted in inadequate deductive criteria of explanation and reduction and as such fail to appreciate the resources a naturalistic, bio-psychological approach has for taking subjectivity seriously. The mind-body or brain-consciousness problem was resolved in favor of a neurophilosophical, inflationist form of naturalism.

Chapter Two then argued that the contemporary mind sciences not only call dualism into question; they also help explain the genesis of robust dualist intuitions in human thought. Such an explanation is part of the burden of proof for naturalists. Naturalists not only have to show that dualism is false; they also have to provide a compelling naturalistic explanation of dualism. The cognitive neuroscience of the self and the cognitive science of religion both figured prominently in this task. Insights from cognitive neuroscience of the self and the cognitive science of religion were drawn upon to explain everyday “out-of-brain” experiences and more extraordinary “out-of-body” experiences. Such experiences are real and vivid enough, but can be explained in entirely naturalistic terms in light of the notions of a transparent phenomenal self model, the notion of the self as a “virtual reality,” and the notion of bodily self representation and multi-sensory integration.

In terms of cognitive science, it was shown that dualism emerges as an evolutionary by-product of other directly adaptive inference systems: folk physics, folk biology, and folk psychology. Dualism emerges as a consequence of the fact that we have two different inference systems for reasoning about social reality (folk psychology) and reasoning about physical reality (folk physics), and from the fact that our mind reading abilities tend to overshoot as a result of the fact that much of our social reasoning occurs offline and that some mental states are easier to imagine being without than other mental states. It was argued that the fact that mind-body dualism comes so naturally to human beings contributes to the genesis of other supernatural concepts like god-concepts. Here dualism couples with another by-product of human cognition: the tendency to over-

attribute agency and design. The promiscuous nature of agency detection and theory of mind leads us to see “faces in the clouds,” to see the biological and non-biological world as designed for a purpose, and to see natural events as symbolic communications.

Chapter Three then turned to address the religious-theological implications of the demise of dualism. I addressed two attempts to integrate religious dualism with the mind sciences. It was argued that Philip Clayton’s attempt to retain a theological dualism while jettisoning dualism as a doctrine about human persons was unsuccessful as was B. Alan Wallace’s more experientially focused defense of religious dualism from a Buddhist perspective.

The Implications of the Naturalistic Conclusions of this Dissertation for the Religion and Science Dialogue

Having recapped the major moves of the dissertation, it is time to take stock of the conclusions thus far and their bearing on issues concerning religion and science. Literature on religion and science tends almost inevitably to understand the relationship between the two in terms of conflict, independence, or integration. In light of this taxonomy, the conclusions of this dissertation seem to be a straightforward example of the conflict model, and indeed they are. However, what is seldom recognized by those who adhere to these various models is that questions concerning the relationship between religion and science are domain specific. This dissertation has argued that claims about metaphysical dualism are an important part of many religious traditions, and if its conclusions thus far are correct, these claims conflict with what contemporary philosophy and science tells us about the nature of mind. The contemporary philosophy and science

of the mind brain supports naturalist or physicalist rather than dualist conclusions. Dualist claims, however, are simply one aspect or dimension of religion, and issues concerning metaphysics and the mind-body problem are merely one point of intersection between the claims of religion and the claims of science. No claims have been made that metaphysical dualism forms the “essence” of religion or that religion collapses tout court with the demise of dualism.

The mind sciences and a naturalism informed by the mind sciences have much more to contribute to the study of religion than a debunking of dualism. Accepting metaphysical naturalism hardly means that there can be no interesting points of intersection between religion and science. The realm of factual and metaphysical claims—such as those involved in the mind-body problem—is an important point of intersection between religion and science, and a point of intersection that is conflictual and not favorable to religion. But the debate concerning metaphysical questions like the mind-body problem is hardly exhaustive of the potential points of intersection between religion and science, and the mind sciences can illuminate religion in a much more significant manner than simply telling us that the dualist claims embedded within religious traditions are likely false. In particular, there are questions about morality and value that religious traditions are deeply concerned with, and that the mind sciences have increasingly shed light on. In this chapter, I turn from metaphysics and the mind-body problem to morality as a point of intersection between religion and science.

Note that this proposal is nearly the exact opposite of the “independence” approach to religion and science advocated by Stephen Jay Gould and others. The issue

of metaphysical truth claims has dominated discussions of religion and science, and Gould was one of the first to recognize the limits of approaches to religion and science framed in terms of total conflict (e.g., New Atheism/Intelligent Design/Creationism) or comprehensive integration (e.g., Christian theologies of science):

...[W]hen we must make sense of the relationship between two disparate subjects (science and religion in this case)—especially when both seem to raise similar questions at the core of our most vital concerns about life and meaning— we assume that one of two extreme solutions must apply: either science and religion must battle to the death, with one victorious and the other defeated; or else they must represent the same quest and can therefore be fully and smoothly integrated into one grand synthesis. But both extreme scenarios work by elimination—either the destruction of one by another, or the merger of both into a large and pliant “whole ball of wax” without sharp edges or incisive points.¹

Gould aptly diagnoses the problems and limits of the conflict and integration approaches, but his own cure is deeply problematic. Gould famously (or notoriously) proposes that religion and science be understood as “non-overlapping magisteria” (NOMA). “The net, or magisterium, of science covers the empirical realm: what is the universe made of (fact) and why does it work that way (theory). The magisterium of religion extends over questions of ultimate meaning and moral value.”² Where conflict and integration strategies tend to view all of religion and science exclusively in terms of one domain—metaphysical truth claims—the NOMA strategy recognizes multiple domains but goes on to assign science and religion exclusive rights to each domain: science gets the domain of facts while religion gets the domain of values. Gould’s recognition of multiple domains is laudable; his approach goes wrong in his parsing of

¹Stephen Jay Gould, *Rocks of Ages: Science and Religion in the Fullness of Life* (New York: Ballantine, 1999), 31.

²*Ibid.*, 6.

issues in terms of non-overlapping magisteria. There is no reason to think that religion cannot make factual claims (whether the claims are true is another matter), further, there is no reason to think that religion is the sole repository of morality and values or that science might not shed interesting light on issues of morality and meaning. I have argued that religious traditions like Buddhism and Christianity have made and continue to make claims about the nature of the human person and the nature of the world, and that these claims intersect in interesting ways with the claims of modern science. I have also argued that there are good reasons to be skeptical of these claims in light of modern science; however, this point of conflict does not exhaust areas of intersection between religion and science, and I suggest that it is precisely in the realm of ethics and morality where a more fruitful exchange might occur.

Points of Intersection between Religion and Science:

Metaphysics and Morals

The approach to religion and science advocated in this dissertation sees both the metaphysical and moral approaches as essential. Claims about metaphysics and morality are intimately related; however, it is important to recognize that fundamentally different questions are at stake when pursuing religion and science from the perspective of metaphysics and when pursuing it from the perspective of morality.

I first want to make some points about the metaphysical intersection and the conclusions that have been reached about it thus far in this dissertation. I have argued that issues concerning metaphysics will inevitably create a conflict between religion and science. The dualist vision found in religious traditions like Buddhism and Christianity

simply does not cohere well with the contemporary philosophy and science of the mind-brain. Many recoil from the conflict stance because it seems inherently dogmatic and polemical, but there is no reason why the honest registering of intellectual disagreement has to occur in a dogmatic and polemical manner.

Does naturalism commit one to a position that is dogmatic, polemical, and anti-religious? It is hard to see why this should necessarily be the case. The worry about dogmatism can be assuaged by recognizing that the form of naturalism defended in this dissertation is a fallibilist form of naturalism. Chapters One, Two, and Three have argued that a modest, fallibilist form of naturalism is the most adequate stance to adopt at this point in time concerning claims about the nature of mind. These are provisional conclusions that could all change on empirical grounds. Empirical evidence and philosophical arguments might develop that lead in dualist rather than naturalist directions. At this point in time, however, this is unlikely.

There is nothing dogmatic about this at all; it is simply following the evidence wherever it leads. As noted above, a more dogmatic kind of naturalism is actually implicit in the seemingly irenic independence/NOMA stance. This stance *does* dogmatically assert that religious claims have/should have no ability to make factual claims about the nature of the world and the nature of the human person. The naturalism defended here in no way entails the claim that religion does not or *should not* make metaphysical claims; it is simply skeptical about the truth value of such dualist claims. That is to say it advocates a certain intellectual stance, but this is far from implying any disrespect for those who hold dualist and supernaturalist positions. Far from dogmatic

table thumping that marks the end of dialogue, the direct naming of conflict provides the basis for a more robust religion and science dialogue.

There is no reason why conflict has to preclude dialogue and even friendship between those holding wildly divergent positions. A model for what I have in mind is suggested by David Hume in his *Dialogues on Natural Religion*. It is significant that Hume uses the genre of dialogue to probe central issues concerning religion. It is particularly significant to note that the deep philosophical and religious differences between Philo and Cleanthes occur in the context of mutual respect and deep and abiding friendship. In his commentary on the *Dialogues*, William Lad Sessions remarks:

[T]hey deeply trust and respect and like each other. They do not merely understand each other's subterfuges and strategies; they have no fear of divulging their deepest hopes and views to one another, even though they do not see things from a single point of view...Theirs is not an arrangement of convenience or domination; it is, to borrow George Fox's great phrase, a society of friends. This society, I believe...is intended to model for Pamphilus an enticing form of life, one that can explore the great issues of religion without erupting into superstition or enthusiasm—a form of life that can enfold deep difference and honest debate within its respect for persons.³

Frank acknowledgement of conflict can in fact contribute to a more robust religion and science dialogue, but I want to suggest that naturalism has a broader significance for religion, theology, and spirituality. Naturalism has a relevance for religion that goes far beyond providing supernaturalists with a friendly sparring partner. One sees this when the point of intersection between religion and science is shifted from metaphysics to morality.

³William Lad Sessions, *Reading Hume's Dialogues: A Veneration for True Religion* (Bloomington & Indianapolis: Indiana University Press, 2002), 230.

Shifting to the issue of morality is not punting on metaphysics, but rather a pragmatic recognition that such discussions only get one so far. The shift allows for collaboration between very different groups around an issue that is of common concern to all—the moral life. Questions of morality are simply of concern to everyone regardless of their religious commitments or lack thereof. Questions about morality thus typically engage a broader audience than issues of religious metaphysics. The shift to morality occurs naturally when one recognizes that dualist metaphysics is no longer tenable. But one does not have to accept metaphysical naturalism in order to appreciate ways in which scientific work in moral psychology might impact issues of morality and religion. In fact, such work may be an important resource for religious traditions. First, religious ethical claims are premised on certain claims regarding the human person and moral psychology. Many of these traditional philosophical-theological claims regarding the human person and moral psychology have been called into question by research in scientific moral psychology. Further, there are massive moral conflicts within religious traditions, and the traditional categories of philosophical and theological analysis have done little to either explain or assuage these points of conflict.

I examine the intersection of mind science, morality, and religion in three sections. As the last three chapters examined the implications of the mind sciences for the metaphysical claims embedded in many religious traditions, the remainder of this chapter examines the implications of the mind sciences—especially scientific moral psychology—for metaethical claims concerning the nature and origin of moral judgment that are imbedded in many religious traditions.

The first section is concerned with questions about the origins of morality, specifically the claim that religion lies at the origin of morality. The issue of the relationship between morality and religion (or morality and God(s)) is ancient and dates back at least to Plato's *Euthyphro* dialogue. Such issues are far from being simply interesting, if esoteric, metaethical questions. Indeed, they cut to the core of the so-called "culture wars" in the United States. At least in the American context, moral pluralism and conflict is linked closely with questions of religious belief. The notion is widespread that religion is the source of morality and that one cannot be moral without a belief in God. Even some who are themselves atheists or agnostics believe that belief in God is essential to keeping the masses moral. On the other hand, certain parts of the "New Atheist" literature suggest not only that religion is false, but that it is very, very bad. Religion is a "delusion" or "spell" that "poisons everything." Such literature gives the suggestion that belief in God may be inherently morally corrupting.

In the past twenty years, a significant research program has been launched by psychologists studying morality that investigates people's ability to distinguish between moral and conventional violations. As it turns out, research on the moral/conventional distinction sheds light on the relationship between religion and morality and the origin of morality. It is argued that such research casts doubt on the claim that religion—especially in the form of divine commands—is the origin of morality and that such research points to an emotional origin of morality. If religion is not the source of morality, how do religion and morality come to link up in the minds of so many people? It is argued that a

commonsense belief in the objectivity of morals provides an important link between morality and religion.

The next section addresses the question of how emotion relates to reason in moral judgments and how both biology and culture contribute to the genesis of moral intuitions. These questions are explored in light of Jonathan Haidt's social intuitionist model of moral judgment. It is argued that this model sheds a great deal of light on issues concerning moral pluralism and conflict. Finally, the implications of this work for normative ethics and religious practices are addressed.

Mind Science, Metaethics, and Religion I:

Divine Command Theory and the Origins of Moral Judgment

Religion and Morality

Today, nearly two centuries after the Enlightenment and its emphasis on the autonomy of ethics, the opinion is still widespread in the United States that religion is the essential origin of morality. It is not difficult to find statements supporting this viewpoint. The nation's presidents seem particularly fond of asserting this link between religion and morality. In a speech to an ecumenical prayer breakfast during the 1984 Republican National Convention, Ronald Reagan would remark: "The truth is, politics and morality are inseparable. And as morality's foundation is religion, religion and politics are necessarily related."⁴ Even those who are not particularly religious seem to share the view that while religion may not be true, it is essential to morality. An important

⁴*New York Times*, August 24, 1984 as quoted in Lawrence S. Cunningham and John Kelsay, *The Sacred Quest: An Invitation to the Study of Religion* (Upper Saddle River, NJ: Pearson/Prentice Hall, 2006), 125.

corollary of this view is that naturalism has devastating consequences for morality and skepticism about whether atheists can, in fact, be moral people. This skepticism about the moral credentials of atheists has been well documented in sociological surveys⁵ and is captured in the following interview with Reagan's successor, George H.W. Bush (arguably the least religious president in the past thirty years).

Q: Surely you recognize the equal citizenship and patriotism of Americans who are atheists?

A: No, I don't know that atheists should be considered as citizens, nor should they be considered patriots. This is one nation under God.⁶

The comments of Reagan and Bush echo a popular sentiment that religion—especially divine commands—are the sine qua non of ethics. This understanding of the origins of morality not only creates antipathy between nonbelievers and believers, it creates antipathy between religious believers who disagree about certain moral issues; both sides ground their moral convictions in the objective will of God. Much then might be gained by empirical investigation into the origins of morality and the relationship between religion and morality. There are, of course, many avenues by which questions concerning the relationship of religion and morality and the origins of morality might be fruitfully studied. Here I investigate these issues drawing upon a fruitful research trajectory developed in scientific moral psychology over the past twenty years that documents people's ability to distinguish between moral and conventional violations.

⁵Penny Edgell, Joseph Gertais and Douglas Hartmann, "Atheists As "Other": Moral Boundaries and Cultural Membership in American Society," *American Sociological Review* 21, no. 2 (2006).

⁶*Free Inquiry* 8, no.4 (1988), 16 as quoted in Walter Sinnott-Armstrong, *Morality without God?* (Oxford: Oxford University Press, 2009), 8.

I first introduce the moral/conventional distinction as defined and developed by Eliot Turiel and others. I then turn to Turiel and Larry Nucci's application of the moral/conventional distinction to religious contexts, arguing that this work sheds much light on vexed question of the relationship between religion and morality. Specifically, it casts doubt on the claim that religion (especially in the form of divine commands) is the source of morality. If religion is not the source of morality what is? Next, drawing on R. James Blair's work on psychopaths and work by Jonathan Haidt and Shaun Nichols on disgust, I argue for an affective basis for the moral/conventional distinction. Drawing in particular on the work of Shaun Nichols, I show how this affective base of morality also generates the belief in the objectivity of morals; following the work of Pascal Boyer and others, I argue that it is primarily though moral objectivity that religion and morality come to be linked. The implications of these findings are then discussed.

Religion, Morality, and the Moral/Conventional Distinction

In the past twenty years, Eliot Turiel and his collaborators have launched an important research program documenting the ability of children to distinguish between morality and social convention. Conventions here are understood as agreed upon behavioral uniformities determined by a particular social system. "Conventions are part of constitutive systems and are shared behaviors (uniformities, rules) whose meanings are defined by the constituted system in which they are embedded."⁷ Moral considerations, on the other hand, "stem from factors intrinsic to actions: consequences such as harm to

⁷E. Turiel, M. Killen, and C. Helwig, "Morality: Its Structure, Functions, and Vagaries," in *The Emergence of Morality in Young Children*, ed. J. Kagan and S. Lamb (Chicago: University of Chicago Press, 1987), 169. As quoted in Shaun Nichols, *Sentimental Rules: On the Natural Foundations of Moral Judgment* (Oxford: Oxford University Press, 2004), 5.

others, violations of rights, effects on the general welfare.”⁸ A number of important findings have been documented in empirical studies concerning the moral/conventional distinction in secular contexts:

- Moral transgressions are viewed as wrong irrespective of the presence of governing rules, while conventional acts are viewed as wrong only if they are in violation of an existing rule or standard.
- Individuals view conventional standards as culturally relative and alterable, while moral prescriptions are viewed as universal and unchangeable.
- The forms of social interaction in the context of moral events differ qualitatively from interactions in the context of conventions. Specifically, it was found that children’s and adults’ responses to events in the moral domain focus on features intrinsic to the acts (e.g. harm, justice), while responses in the context of conventions focus on aspects of the social order (e.g. rules, regulations, normative expectations).
- Individuals view moral transgressions as more serious than violations of convention.
- Prosocial moral acts are viewed as better or more positive than adherence to conventions.⁹

Larry Nucci extended Turiel’s work on the moral/conventional distinction into the religious sphere. As Nucci and Turiel note, the religious sphere raises interesting issues concerning the moral/conventional distinction not present in secular contexts:

Religious frameworks provide a useful context for further research into children’s moral and social concepts. In the first place, some rules specific to particular religions, which by the criteria used in the domain model would be classified as nonmoral, are usually treated as important and binding on their members. Secondly, “moral” rules are often closely linked to religious authority. It may be, therefore, that religious people do

⁸Larry P. Nucci, "Children's Conceptions of Morality, Societal Convention, and Religious Prescription," in *Moral Dilemmas: Philosophical and Psychological Issues in the Development of Moral Reasoning*, ed. Carol Gibb Harding (Chicago: Precedent Publishing, 1985), 139.

⁹*Ibid.*, 141.

not distinguish between moral and non-moral religious rules. That is, they may view both types of rules as prescriptive and universally binding.¹⁰

Nucci and Turiel were concerned to see whether classic moral violations would be distinguished from other religious behaviors, and, most important for our purposes here, whether moral judgments would be made in a manner contingent upon religious considerations (e.g., God's Word as revealed in the Bible, etc.).

In one study, sixty-four Amish-Mennonite children were asked various questions about various moral (e.g., stealing, hitting, slander, damaging property) and non-moral issues (e.g. day of worship, work on Sabbath, baptism, head coverings, interfaith marriage, and premarital sex).¹¹ The Amish-Mennonite children were asked three sets of questions. The question sets concerned rule alterability, act generalizability, and "God's word contingency":

- (1) Rule Alterability: "Suppose all of the members of the congregation and the ministers agreed to alter/eliminate the rule about [the act], would it be wrong or all right for them to do that? Why/why not?"¹²
- (2) Act Generalizability: "Suppose that in another religion they don't have a rule about [the act], would it be wrong or all right for them to [engage in the act] in that case? Why/why not?"¹³
- (3) "God's Word Contingency": "Suppose there was nothing in the Bible about [the act], God hadn't said anything about [the act], would it be wrong or all right for a Christian to [engage in the act] in that case? Why/why not?"¹⁴

¹⁰Larry Nucci and Elliot Turiel, "God's Word, Religious Rules, and Their Relation to Christian and Jewish Children's Concepts of Morality," *Child Development* 64, no. 5 (1993): 1476.

¹¹Ibid.

¹²Ibid., 1478.

¹³Ibid.

¹⁴Ibid.

Nucci and Turiel found strong distinctions between the moral and non-moral domains in all three question sets. With regard to alterability, the great majority of Amish subjects judged that it would be wrong for religious leaders to alter rules concerning stealing/hitting/slander/property damage [59/60/59/61 (N = 64)].¹⁵ This was the only set of moral questions where subjects most commonly made appeal to God's law as a justification for their responses. God's law justifications were cited most frequently (.58), followed by justifications that appealed to intrinsic features of the act (harm/fairness/obligation) (.38), social system (.01), and other justifications (.03).¹⁶ While judgments regarding the alterability of non-moral religious rules were significantly lower in certain cases (e.g., only 27 subjects regarded the rule prohibiting interfaith marriage as unalterable by religious authorities), other non-moral rules were regarded as unalterable.¹⁷ Fifty-nine respondents regarded work on the Sabbath as unalterable by religious authorities. Here again, God's law was the most commonly cited justification (.57), followed by social system (.34) and other (.11) justifications.¹⁸

Turiel and Nucci's findings become more interesting with regard to generalizability and the God's word contingency questions. Numbers were close to ceiling regarding the question of whether it would be wrong for members of another religion to steal/hit/slander/damage property even if that religion did not have a rule

¹⁵Ibid., 1480.

¹⁶Ibid., 1481.

¹⁷Ibid.

¹⁸Ibid.

about that action [64/58/62/57 (N = 64)].¹⁹ In terms of the non-moral religious rules, despite their appeal to God's law as a justification for the inalterability of these rules by religious authorities, numbers were quite low regarding the generalizability of non-moral religious rules. What is particularly interesting is the shift in frequencies of justifications. In the moral domain with regard to generalization, intrinsic features justifications were by far the most common (.77) with God's law justifications accounting for a small proportion of justifications (.13).²⁰ In the non-moral domain regarding generalizability, justifications were primarily based on social system (.64), followed by appeal to God's law (.23) and other justifications (.09).²¹ With regard to the God's word contingency question, the great majority of subjects regarding stealing/hitting/slander/property damage as wrong even if God had not made a rule about them [54/56/54/58 (N = 64)].²² Here justification overwhelmingly focused on intrinsic features (.94) versus God's law (.03) and other (.03) justifications.²³ In short, the great majority of Amish subjects regarded moral violations not to be contingent upon God's will. As Nucci and Turiel comment, "Nearly all subjects judged that actions in the moral domain constitute transgressions even in the absence of a command from God opposing the act because of the intrinsic consequences these acts have upon others."²⁴ On the other hand, non-moral

¹⁹Ibid., 1480.

²⁰Ibid., 1481.

²¹Ibid.

²²Ibid., 1480.

²³Ibid., 1481.

²⁴Ibid., 1481-1482.

religious violations were regarded as strongly contingent upon God's will. No subjects regarded rules concerning day of worship, work on Sabbath, baptism, women's head coverings, and interfaith marriage to be wrong in absence of a specific divine command.²⁵

The relationship between God's word and morality was probed further by Turiel and Nucci in an experiment with Dutch Calvinist children. Sixty-four Dutch Calvinist children were asked the following question based on the Euthyphro Dilemma:

- (4) "Open Question": Suppose God had commanded [written in the Bible] that Christians should steal. Would it then be right for a Christian to steal?" "Do you think God would make a commandment saying that Christians should steal?"²⁶

Most children maintained that God's command to steal would not make it right to steal (11 of 16 at 10 and 12 years; 13 of 16 at 14 and 16 years).²⁷ Nucci has replicated these findings across a number of religious traditions.²⁸ Even children from very religious backgrounds do not see morality exclusively in terms of their religious commitments. As Nucci and Turiel write:

The results from Study 1 provide evidence that Amish-Mennonite children's conceptions of morality are not reducible to their knowledge of or adherence to religious rules. Consonant with prior research on Catholic subjects (Nucci 1982), the Amish-Mennonite and Dutch Reform Calvinist children evaluated moral issues in terms of justice and welfare considerations rather than precepts of the Bible or positions taken by religious authorities. As did the Catholics, the Amish-Mennonite children

²⁵Ibid.

²⁶Ibid., 1478.

²⁷Ibid., 1483.

²⁸Ibid.

generalized moral issues, and viewed moral rules as unalterable by religious authorities. They also viewed the status of moral transgressions as noncontingent on God's word. Furthermore, most of the Dutch Reformed Calvinist children responded that God's command would not make stealing right. These findings indicate that children from these groups maintain a distinct moral position based on justice and welfare criteria from which they apprehend the moral aspects of the Christian God."²⁹

The Origin of the Moral/Conventional Distinction

The moral/conventional distinction is interesting because it shows that moral judgments are arrived at independently of authority including religious authority and even God's authority. If morality does not come from religion, where does it come from? Alongside theological and religious conceptions of morality and intimately related to them are conceptions of morality that understand moral judgment to be rooted in reason. Rationalism has exerted a massive influence on Western thought from philosophers like Plato, Aquinas, and Kant to moral psychologists like Piaget and Kohlberg. This model has been called into question by research in both neuroscience and psychology. This research suggests that it is affect rather than reason that drives many moral judgments and actions.

The Affective Origins of Morality:

Psychopaths and the Moral/Conventional Distinction

The moral/conventional distinction has been studied in a number of interesting contexts. Some of the most interesting findings concerning the moral/conventional

²⁹Nucci, "God's Word, Religious Rules, and Their Relation to Christian and Jewish Children's Concepts of Morality."

distinction come from R. James Blair's work with psychopaths and autistics.³⁰ Blair's research has documented that criminal psychopaths and children with psychopathic tendencies fail to treat moral violations and conventional violations as significantly different. Blair's studies show that psychopathic adults and children fail to draw an adequate moral/conventional distinction and also tended to ignore considerations about the welfare of victims when justifying their responses.³¹ Children with psychopathic tendencies tended to regard moral violations as authority dependent.³² Psychopathic individuals seem to have a sense of right and wrong, but fail to distinguish between different *kinds* of wrongs or violations.³³ Significantly, psychopaths do not seem to suffer from any rational defects. They do not differ from others on measures of intelligence. Perhaps psychopaths suffer from the loss of a critical faculty essential for moral reasoning. For example, perspective taking and theory of mind suggests itself. Such perspective-taking abilities have been an important focus in rationalist accounts of moral psychology. Testing this hypothesis, Blair found that psychopaths scored *slightly better*

³⁰The philosophical implications of Blair's work are discussed in Shaun Nichols, *Sentimental Rules: On the Natural Foundations of Moral Judgment* and Jesse Prinz, *The Emotional Construction of Morals*. The following section is indebted to the work of both Nichols and Prinz.

³¹R. Blair, "A Cognitive Developmental Approach to Morality: Investigating the Psychopath," *Cognition* 57, no. 1 (1995); R. Blair, "Moral Reasoning and the Child with Psychopathic Tendencies," *Personality and Individual Differences* 22, no. 5 (1997).

³²Blair, "Moral Reasoning and the Child with Psychopathic Tendencies."

³³In his discussion of psychopathy, Shaun Nichols illustrates this with a quotation from the presumed psychopath Ted Bundy. "It was almost as if he [I] said it was wrong for all these things to happen. "It is wrong for me to jaywalk. It is wrong to rob a bank. It is wrong to break into other people's houses. It is wrong for me to drive without a driver's license. It is wrong not to pay your parking tickets. It is wrong not to vote in elections. It is wrong to intentionally embarrass people" (Nichols, 2004, 65).

than a control group on measures of perspective taking and theory of mind.³⁴ Further, autistic individuals are known to have seriously impaired theory of mind and perspective taking abilities. Significantly however they do not manifest the difficulties that psychopaths do in drawing the moral/conventional distinction.³⁵ In short, psychopaths seem to suffer from no rational deficit at all; however, psychopaths do show significant differences in emotion and affect. Blair and his colleagues have tested emotional response by showing psychopaths and a wide number of control groups pictures of both threatening and distressed faces.³⁶ All of the control groups showed significantly heightened physiological responses to both the threatening and distressed faces; however, the psychopaths showed response to only the threatening faces. A number of other significant affective and emotional deficits have been documented in psychopathy.³⁷ This research suggests that moral concepts are *essentially* emotion-involving concepts. Jesse Prinz makes this point well:

If moral properties were not essentially emotion-involving, then there should be a way of drawing the moral/conventional distinction without appeal to emotions. Psychopaths should be able to learn the difference. After all, psychopaths generally have intelligence quotients within the normal range. Some are extremely bright. They are often articulate and cunning. They have a great interest in learning to distinguish right from

³⁴R. Blair, C. Sellars, I. Strickland, F. Clark, A. Williams, M. Smith, and L. Jones, "Theory of Mind in the Psychopath," *Journal of Forensic Psychiatry* 7, no. 1 (1996).

³⁵R. Blair, "Brief Report: Morality in the Autistic Child," *Journal of Autism and Developmental Disorders* 26, no. 5 (1996).

³⁶R. Blair, "Responsiveness to Distress Cues in the Child with Psychopathic Tendencies," *Personality and Individual Differences* 27, no. 1 (1999).

³⁷See, for examples, the articles on psychopathy in Walter Sinnott-Armstrong, ed. *Moral Psychology: Volume 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development* (Cambridge, MA: MIT Press, 2008).

wrong. Most psychopaths have long histories of misconduct. . . . These histories put them into contact with people who take special care in helping them grasp morality. They often encounter concerned parents, teachers, lawyers, and law enforcers. Given this combination of exposure and motivation to learn, psychopaths should be more likely than others to develop ways of reliably identifying moral properties. The fact that they fail to master the moral/conventional distinction suggests that there may be no way to draw that distinction without adverting to or experiencing emotional responses. The distinction may be emotional to the core. Right and wrongness may be constituted by emotional reactions in us. Subtract these reactions and the distinction becomes as invisible as the color spectrum is to the blind. Psychopaths can carefully monitor the moral judgments of healthy individuals. They can sort familiar examples of good conduct into one bin and bad conduct into another, but they lack insight into the very essence of this division.³⁸

The Affective Origins of Morality: The Moral Psychology of Disgust

Studies of disgust also shed significant light on the moral/conventional distinction and the role of emotion in moral judgment. Jonathan Haidt has done particularly interesting cross cultural work on people's reactions to harmless taboo violations. Are such actions subject to moral condemnation, and if so, why? Haidt's subjects varied in nationality, socioeconomic status (SES), and age. Subjects were presented with a number of harmless taboo violations: a family eats its pet dog after it was hit by a car, a woman cuts up a national flag and uses the strips to clean her toilet, a man uses a chicken carcass for masturbation and afterwards cooks and eats the carcass.³⁹ Haidt and his colleagues found that a significant number of groups condemned these actions morally. Haidt was particularly concerned with the issue of why these actions were condemned. Perhaps the actions were understood to have *harmful* consequences even though they really did not.

³⁸Jesse Prinz, *The Emotional Construction of Morals* (Oxford: Oxford University Press, 2007).

³⁹Jonathan Haidt, Silvia Helena Koller, and Maria G. Dias, "Affect, Culture, and Morality, or Is It Wrong to Eat Your Dog?," *Journal of Personality and Social Psychology* 65, no. 4 (1993).

Or perhaps they were morally condemned because one was *bothered* by seeing such an action. The study found that among the groups that morally condemned these actions, the *bother* probe was a better predictor of negative judgment than the *harm* probe. Such a finding fits well with an understanding of moral judgments driven by emotional intuitions. Differences between the “moralizing” and “permissive” groups also supports the thesis that emotional judgment is the cognitive default that sometimes can be overridden by more explicitly cognitive processes.⁴⁰ High SES groups—college students—in both the United States and Brazil tended not to judge the actions as moral. People from less Westernized cities tended to be more condemning, and children in both the United States and Brazil were more condemning than adults.⁴¹ As Joshua Greene reflects:

These three findings make sense in light of the model of moral judgment we have been developing, according to which intuitive emotional responses drive prepotent moral intuitions while “cognitive” control processes sometimes rein them in. Education is to a large extent the development of one’s “cognitive” capacities, learning to think in ways that are abstract, effortful, and often either nonintuitive or counterintuitive. The westernization factor is closely related. While westerners may not be any more “cognitively” developed” than members of others cultures, the western tradition takes what is, from an anthropological perspective, a peculiarly “cognitive” approach to morality. Westerners are more likely than members of other cultures to argue for and justify their moral beliefs and values in abstract terms...Moreover, western culture tends to be more pluralistic than other cultures, explicitly valuing multiple perspectives and an intellectual awareness that alternative perspectives exist. Finally, the capacity for “cognitive control” continues to develop through

⁴⁰Joshua D. Greene, “The Secret Joke of Kant’s Soul,” in *Moral Psychology: Volume 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, ed. Walter Sinnott-Armstrong (Cambridge, MA: MIT Press, 2008).

⁴¹Jonathan Haidt, Silvia Helena Koller, and Maria G. Dias, “Affect, Culture, and Morality, or Is It Wrong to Eat Your Dog?”

adolescence....Children like adults, are very good at feeling emotions such as anger, sympathy, and disgust, but unlike adults they are not very good at controlling their behavior when experiencing such feeling.⁴²

Haidt and his colleagues have also manipulated disgust through hypnotic suggestion and the alternation of environmental factors in order to assess the effect of such manipulations on moral judgment.⁴³ In one set of experiments, half of a group of highly-hypnotizable subjects was given the post-hypnotic suggestion to feel disgust upon reading the word *take*. The other half was given the suggestion to feel disgust upon reading the word *often*. The “*take*” group received three moral violation vignettes containing the word “*take*,” three neutral buffer stories that contained neither the word “*take*” or a moral violation, and three moral violation stories containing the word “*often*.”⁴⁴ Likewise the “*often*” group received three moral violation vignettes containing the word “*often*,” three neutral buffer stories, and three moral violation stories containing the word “*take*.”⁴⁵ The moral violations included accounts of a sexual relationship between second cousins, a man eating his already dead dog, a congressman taking bribes, an ambulance-chasing lawyer, a shoplifter, and a student stealing library books.⁴⁶ Participants evaluated both “how disgusting” and “how morally wrong” the behavior in the vignette by picking slash markers on a line ranging from *not at all disgusting/*

⁴²Greene, 56.

⁴³Thalia Wheatley, and Jonathan Haidt, "Hypnotic Disgust Makes Moral Judgment More Severe," *Psychological Science* 16, no. 10 (2005).

⁴⁴Ibid., 781.

⁴⁵Ibid.

⁴⁶Ibid., 780-781.

extremely disgusting and *not at all morally wrong/extremely morally wrong*. Wheatley and Haidt found that moral transgressions were viewed as more disgusting when their hypnotic disgust word was embedded in the vignette than when it was absent.⁴⁷ Further, they found that the presence of the disgust word increased the severity of moral judgment.⁴⁸

In one of the studies conducted, a story was included where there was no violation at all: “Dan is a student council representative at his school. This semester he is in charge of scheduling discussions of academic issues. He [tries to take up] [often picks] topics that appeal to both professors and students in order to stimulate discussion.”⁴⁹ Haidt and his colleagues predicted that with no violations of any kind subjects would have to override their feelings of disgust and most did. However, one third of all subjects who encountered their disgust word in the story still rated Dan’s actions as somewhat morally wrong. In another set of experiments, Haidt and colleagues manipulated disgust environmentally.⁵⁰ Subjects were asked to make moral judgments at either a clean desk or a dirty desk filled with fast food wrappers and dirty tissues. Those at the dirty desk made moral judgments more severe for individuals scoring in the upper half of a scale measuring “private body consciousness.”⁵¹

⁴⁷Ibid.

⁴⁸Ibid.

⁴⁹Ibid., 782.

⁵⁰S. Schnall, J. Haidt, and G. Clore, "Irrelevant Disgust Makes Moral Judgment More Severe, for Those Who Listen to Their Bodies" (University of Virginia, 2004).

⁵¹Ibid.

Haidt's studies of disgust provide further evidence for the emotional origins of morality. They also provide evidence that the scope of morality is much broader than the harm/justice based norms that figure so centrally in modern Western conceptions of morality. Shaun Nichols has also done interesting experimental work on disgust, particularly in relation to the moral/conventional distinction. In Nichols's "sentimental rules" explanation of the moral/conventional distinction, the basis for the moral/conventional distinction lies in the fact that moral norms are backed by affect while conventional norms are not. Norms backed by affect will be regarded as more serious and less authority contingent than norms that are not affectively backed. Importantly, like Haidt and in contrast to Blair, Nichols understands affect-backed norms as broader in scope than harm-based norms. Nichols tested his sentimental rules account of the morality by replacing the moral/conventional distinction with a disgusting/conventional distinction. Subjects were presented with moral violations, disgusting violations, and conventional violations. Nichols found that subjects tended to treat disgusting violations as less permissible, more serious, and less authority contingent than the affect-neutral conventional violations.⁵² Nichols also tested the disgusting/conventional distinction in a manner analogous to Blair's application of the moral/conventional distinction to psychopaths. There an affective defect impedes psychopath's abilities to distinguish between moral and conventional violations. Nichols wanted to see if different sensitivity to disgust would produce different judgments concerning the seriousness and authority

⁵²Shaun Nichols, "Norms with Feeling: Towards a Psychological Account of Moral Judgment," *Cognition* 84 (2002).

contingency of a disgusting violation. Nichols found that while there was no statistically significant difference between low and high disgust subjects on the permissibility question, the low disgust group regarding the disgusting violation as much less serious and also as more authority contingent.⁵³

Commonsense Moral Objectivism and the Relationship Between Religion and Morality

As a philosopher, Nichols has been particularly concerned to address the implications of the findings of scientific moral psychology about the centrality of emotion in moral judgment to issues about moral objectivity. “Moral objectivity” is a notoriously multivalent term. *Epistemic* moral objectivity is related to the notion of impartiality and freedom from bias. This is not the notion of objectivity under discussion here. Here we are discussing metaphysical or ontological moral objectivity: roughly, the notion that certain actions are right or wrong, good or bad in and of themselves. Nichols, along with a number of other contemporary moral psychologists and moral philosophers, argues that these experimental findings support Humean non-objectivism.

The basic thesis of moral objectivism...is that the moral status of an action is constituted by the action “as it is in itself” and accordingly, true moral judgments are nonrelativistically true. Humean sentimentalism is often regarded as directly opposed to moral objectivity in this sense.... According to Humean sentimentalism, morality hangs entirely on the particular emotional repertoire of humans. Here is one place where Hume seems to be making this point: “If we can depend upon any principle, which we learn from philosophy, this, I think, may be considered as certain and undoubted, that there is nothing, in itself, valuable or despicable, desirable or hateful, beautiful or deformed; but that these attributes arise from the particular constitution and fabric of human sentiment and affection” (1742 [1987], 162). Norman Kemp Smith

⁵³Ibid.

paraphrases Hume's view using much the same language: "Moral judgments, in marking out the good and the evil, have their source not in the eternal nature of any independent reality, but solely in the particular fabric and constitution of the human species" ... In brief, moral judgment depends on certain emotions, and these emotions themselves are rationally arbitrary, so moral judgment, as we know it, is not objective. No action is wrong simpliciter. At best, an action is wrong relative to a population – the population of individuals that share a certain emotional repertoire. A more careful way to frame the Humean conclusion here is that the commonsense commitment to objectivity is unwarranted. Given the emotional basis of moral judgment, we are not justified in our belief that morality is objective.⁵⁴

Nichols argues objectivist ontology of morals is philosophically suspect but also has done initial empirical research that indicates that a belief in moral objectivity may be the default stance of commonsense metaethics. Experiments on the moral/conventional distinction provide evidence for a certain kind of objectivity. We can call this kind of objectivity non-conventional objectivity. Indeed, some of the literature on the moral/conventional distinction seems to regard evidence for non-conventional objectivity as evidence for objectivity simpliciter. However, as Nichols notes, non-conventional objectivity does not exhaust the kinds of metaphysical objectivity. A key philosophical notion of metaphysical objectivity turns on the notion of response-dependence. A commitment to moral objectivity is seen as a commitment to the view that moral properties are not response dependent.

Nichols conducted an experiment to test for a commitment to moral objectivity in this broader sense. Nichols and Trisha Fold-Bennett prepared vignettes concerning gustatory ("yummy"), moral ("good"), and aesthetic ("beautiful") properties and then asked children questions about generalizability and preference dependence.

⁵⁴Nichols, *Sentimental Rules: On the Natural Foundations of Moral Judgment*, 184.

GENERALIZABILITY

Now, think about a long time ago, before there were any people. There were still Grapes, just like the grapes now. Way back then, before there were people, were grapes yummy?

PREFERENCE DEPENDENCE

You know, I think grapes are yummy too. Some people don't like grapes. They don't think grapes are yummy. Would you say that grapes are yummy. Would you say that grapes are yummy *for some people* or that they're yummy *for real*?

GENERALIZABILITY

Now, think about a long time ago, before there were any people. There were monkeys back then too. Way back then, before there were people, when one monkey helped another monkey that got hurt, was that good?

PREFERENCE DEPENDENCE

You know, I think it was good for the monkey to help the other monkey. Some people don't like it when monkeys help each other when they're hurt. They don't think it's good when monkeys do that. Would you say that when one monkey helps a hurt monkey that is good for *some people* or that it's good *for real*?

GENERALIZABILITY

Now, think about a time long ago, before there were any people. There were roses back then too. Way back then, before there were people, were roses beautiful?

PREFERENCE DEPENDENCE

You know, I think roses are beautiful too. Some people don't like roses. They don't think roses are beautiful. Would you say that roses are beautiful *for some people* or that they're beautiful *for real*?⁵⁵

Significantly, Nichols found that while children regarded properties like “yummy” and “fun” as response dependent, children tended not to regard moral and aesthetic properties as response dependent. There is thus initial experimental evidence that objectivism is the default stance of folk metaethics. The commitment to an

⁵⁵Ibid., 174. See Shaun Nichols and Trisha Folds-Bennett, "Are Children Moral Objectivists? Children's Judgments About Moral and Response-Dependent Properties," *Cognition* 90, no. 2 (2003).

objectivist metaphysics of morals may be similar to the commitment to metaphysical dualism: a false position that is nonetheless the cognitive default.

What is interesting is that this evidence about people's understanding of moral objectivity gives us a plausible point of contact between religion and morality. Religious subjects regarded moral violations as non-contingent upon God's will. Thus, following Nucci's experimental evidence across a variety of religious traditions, it is unlikely that religion (i.e., divine commands) is the *origin* of the belief in moral objectivity. The origin of the belief in moral objectivity plausibly lies in the affectively charged nature of our moral beliefs. The cognitive mechanisms giving rise to religion and morality are likely separate, but they become linked in important ways. One important link is moral objectivity. On the one hand, moral objectivity makes religious beliefs more salient. On the other hand, a religious/theological context helps to sustain commitment to moral objectivity. The notion that religious beliefs are in some ways parasitic on moral beliefs and that morality makes religious representations more salient has been underscored by Pascal Boyer.

Moral intuitions are part of our mental dispositions for social interaction. But why are they connected to gods and spirits and ancestors? To understand how such beings fit with moral understandings, consider two facts that I mentioned earlier. First, our moral intuitions suggest to us, from the youngest age, that behaviors are right or wrong *by themselves*, not depending on who considers them, or from what point of view. Second, gods and spirits and ancestors are generally considered *interested parties* in moral choices and moral judgments, rather than providers of codes or rules. These two facts are just two aspects of the same mental processes....

So concepts of gods and spirits are made more *relevant* by the organization of our moral understandings, which by themselves do not especially require any gods or spirits. What I mean by relevant is that the concepts, once put in this moral context, are easy to represent and that

they generate many new inferences. For instance, most people feel some guilt when acting in a way that they suspect is immoral. That is, whatever their self-serving justifications, they may have the intuition that an agent with full description of the situation would still classify it as wrong. Now thinking of this intuition as “what the ancestors think of what I did” or “how God feels about what I did” provides an easy way of representing what is otherwise extremely vague. That is, most of our moral intuitions are clear but their origin escapes us, because it lies in mental processes that we cannot consciously access. Seeing these intuitions as someone’s viewpoint is a simpler way of understanding why we have these intuitions. But this requires the concept of an agent with full access to strategic information....

To sum up, then: Our evolution as a species of cooperators is sufficient to explain the actual psychology of moral reasoning, the way children and adults represent moral dimensions of action. But then this requires no special concept of religious agent, no special code, no models to follow. However, once you have concepts of supernatural agents with strategic information, these are made more salient and relevant by the fact that you can easily insert them in moral reasoning that would be there in any case. To some extent religious concepts are *parasitic* upon moral intuitions.⁵⁶

Moral intuitions help make religious beliefs more relevant and salient. While not the *source* or *origin* of commitment to moral objectivity, it may also be that this religious/theological context helps to *preserve* a commitment to moral objectivity. Shaun Nichols explores this idea:

Although religious considerations do not seem to explain the acquisition of the belief in objectivity, religious beliefs might play an important role in shoring up the belief in objectivity. The relationship between religion and morality is complicated, of course. Perhaps most importantly, most people regard voluntarism (or Divine Command Theory) as implausible—that is, at least many adults think that right and wrong are not decided by God, and Nucci’s evidence suggests that even Amish teenagers share this view. Nevertheless, it is undoubtedly part of the doctrine surrounding many religions, including those in the Abrahamic tradition, that morality is objective. The moral truths, if not created by God, are certainly known by

⁵⁶Pascal Boyer, *Religion Explained: The Evolutionary Origins of Religious Thought* (New York: Basic Books, 2001), 189-191.

God. This would provide an important preservative role for the belief in objectivity.⁵⁷

The previous several sections have reached some tentative conclusions regarding the origins of morality and the relationship of religion to morality: (1) Religion, especially in the form of divine commands, is not the likely origin of morality; (2) Morality likely has its origins in human emotions; (3) As such, moral properties are response-dependent properties; (4) Morality is non-objective but morality is commonly regarded as objective; (5) Moral objectivity is one possible way in which religion and morality link up.

These findings cast interesting light on recent debates between atheists and theists, particularly on issues concerning divine command ethics and theological accounts of moral objectivity. While there are many ways to formulate a religious or theological ethics, divine command theory enjoys widespread popularity, and atheist critics of religion typically make divine command theory the object of their attack. But it may be that both theists and atheists have exaggerated the importance and significance of divine command ethics.

Recall that in Turiel and Nucci's research, even very religious children regarded canonical moral violations to be wrong in the absence of a specific command from God, and even very religious children thought that a divine command could not render a moral transgression right. There is thus initial empirical evidence that even very religious children tend to reach moral judgments in a manner that is largely autonomous from religion. This has important descriptive and normative implications. Descriptively, it

⁵⁷Nichols, *Sentimental Rules: On the Natural Foundations of Moral Judgment*, 183.

indicates that one must look elsewhere for an account of the origins of morality and the sources of our morality judgments. From a more normative standpoint, it casts doubt on a naïve equation of belief in God with moral goodness and the absence of such a belief with moral badness. If divine commands play such a negligible role in the moral judgments of very religious children, one could hardly see them as the *sine qua non* of morality. If divine command ethics has such little empirical traction with regard to the moral judgments of members of religious communities, it can hardly be regarded as essential to ethics or as a prescription for societal health. There hardly seems to be a justification for regarding those who approach ethics in a manner autonomous from divine commands as morally suspect, unless one is prepared to regard Amish and Dutch Calvinist children as morally suspect or at least deeply mistaken in the way they go about making moral judgments. It may be that divine command ethics is a form of theological correctness. A religious belief explicitly endorsed by members of certain religious communities that actually plays a negligible role in actual concrete moral judgments.

Interestingly, this research cuts both ways. Atheists also tend to place a great deal of emphasis on divine command ethics, focusing on its negative aspects; but moral judgments based on divine commands may be less common than atheists think, even in traditions (e.g., Dutch Calvinism) that emphasize theological voluntarism. Indeed atheists and secularists might be heartened by the fact that most Dutch Calvinist children thought that stealing would still be wrong even if God had commanded/written in the Bible that they should steal. Atheists and theists may, in fact, share more in common in terms of their moral judgments than either side would initially think. There is some initial

evidence pointing in this direction. In the Harvard Moral Sense Test, Marc Hauser and his colleagues have analyzed the responses of two hundred thousand people representing diverse cultural and religious backgrounds. Subjects were presented with a contrast between a harmful action and a significant benefit in terms of lives saved. Responses to the dilemmas are remarkably similar across cultural and religious differences. As Hauser notes, "More specifically, in dozens of dilemmas, and with thousands of subjects, the pattern of moral judgments delivered by subjects from a religious background do not differ from those who are atheists."⁵⁸ Religion no doubt interacts with and shapes morality in interesting ways, but it is doubtful that religion is the origin of morality. The traditional worry that "if God is dead, then everything is permitted" seems unfounded. Interestingly, there may be a group for whom "all is permitted," but what appears to be absent in this group is not belief in God or even intelligence or a crucial cognitive capacity but rather the absence of a crucial emotional repertoire. Work on psychopaths and disgust underscores the centrality of emotion in moral judgment. This makes moral truths anthropocentric truths and moral properties response dependent properties. However, human beings tend to be objectivists about morality, and this commonsense belief in moral objectivity may provide a plausible point of contact between religion and morality. Moral objectivism is a false but extremely common metaethical thesis that links up with religion in important ways. Religious belief and commitment to moral objectivity seem to go hand in hand. Goodness or rather "The Good" is located entirely outside the human person in a transcendental realm which human beings can access. There are many

⁵⁸ Ilkka Pyysiainen, and Marc Hauser, "The Origins of Religion: Evolved Adaptation or by-Product?," *Trends in Cognitive Sciences* 14, no. 3 (2009): 106.

who believe that a theologically grounded moral objectivism is essential for normative ethics. Thus Linda Zagzebski writes:

Isn't the theist as much faced with the problem of moral pluralism as is the non-theist? In response, it seems to me that it is certainly true that the theist is faced with a certain amount of skepticism about his own particular moral judgments. Theism does not guarantee the possession of moral truth to the believer. However, the theist has a better way of dealing with the problem of moral pluralism in two ways. First, the theist has another source of moral knowledge in divine revelation and the teachings of the Church.... This is not to suggest that the theist has no problems with doubt about the interpretation of God's will. Such a suggestion would be naïve. But again, the Christian concept of providence provides confidence that these problems are resolvable.⁵⁹

Theological objectivism seems to provide a neat solution to the problem of moral pluralism. However, there exist multiple religious communities and even within religious communities— say the Anglican Communion or the Roman Catholic Church—there is robust disagreement on moral issues. These moral disagreements simply mirror those of secular society. Contrary to Zagzebski's claim that the theist has a better way of dealing with the challenge of moral pluralism, appeal to a transcendent source of moral objectivity seems to make moral disagreement even more intractable, as claims about the transcendent have proved difficult if not impossible to adjudicate. Peter Byrne aptly describes these difficulties:

There is more than one claim to be the source of true revelation and the true Church. Where there is agreement on these things, we can find disagreement over how they should be interpreted. The fact of moral diversity itself encourages disagreements on what is religious authority and how it is to be interpreted. Finally, if Zagzebski is right, we should expect to find that those in receipt of the true revelation on moral matters

⁵⁹Linda Zagzebski, "Does Ethics Need God?," *Faith and Philosophy: Journal of the Society of Christian Philosophers* 4, no. 3 (1987), 302.

and those possessing the teachings of the true Church in this area should stand out as having greater moral knowledge than those less fortunate. These folk should stand out as the ones able to rise above disagreements and settle them. Their views are not subject to the relativities and uncertainties others are enmired in. Is there such a group of moral agents? Who are its members?⁶⁰

There have yet to be any compelling answers to the questions raised by Byrne. There are thus numerous problems with a theological account of moral objectivity. In terms of metaphysics, theological objectivism assumes the truth of theism, and as noted in the previous chapters of this dissertation, there are many reasons to be skeptical about the truth of theism. Secondly, there is a massive amount of empirical evidence implicating the emotions in moral judgment. This research calls into question the plausibility of moral properties like “good” and “bad” “right” and “wrong” existing independently of homo sapiens. In terms of epistemology, there is no sort of consensus on the moral issues the way there is with other objective properties (e.g., the roundness of the Earth). As Byrne’s quote illustrates, this lack of consensus exists both *between* even *within* religious communities. Transcendental claims about moral objectivity are problematic on both metaphysical and epistemological grounds, but such claims also might be problematic from a moral standpoint. Moral objectivism contributes to the intractability of moral disputes, especially when couched in theological terms. It may enhance a cognitive bias that has been termed “direct realism”: the well-documented psychological tendency to regard oneself or one’s group as right and others as biased. Jonathan Haidt highlights the ethical dangers of such a stance:

⁶⁰Peter Byrne, *The Moral Interpretation of Religion* (Grand Rapids: Eerdmans, 1998), 80.

If I could nominate one candidate for ‘biggest obstacle to world peace and social harmony,’ it would be naïve realism because it is so easily ratcheted up from the individual to the group level: My group is right because they see things as they are. Those who disagree are obviously biased by their religion, their ideology, or their self-interest.⁶¹

One might be well advised to abandon or at least bracket claims about moral objectivity and to analyze more deeply the sources of moral pluralism and conflict. The next sections will provide such an analysis focusing on the relationship of reason and emotion and biology and culture in morality.

Mind Science, Metaethics, and Religion II: Jonathan Haidt and Joshua Greene on the Relationship of Emotion and Reason in Moral Judgment

Work on the centrality of emotion in morality fits nicely with evidence from contemporary psychology concerning the centrality of automatic intuition in human thought.

Intuitions are the judgments, solutions, and ideas that pop into consciousness without our being aware of the mental processes that led to them...Moral intuitions are a subclass of intuitions, in which feelings of approval and disapproval pop into awareness as we see and hear about something someone did or as we consider choices for ourselves.⁶²

Work on the intuitions and automatic processing has led to a rediscovery of the unconscious in contemporary scientific, experimental psychology. This is not the psychoanalytic unconscious of Freud and his followers but rather an adaptive

⁶¹Jonathan Haidt, *The Happiness Hypothesis: Finding Modern Truth in Ancient Wisdom* (New York: Basic Books, 2006), 71.

⁶²Jonathan Haidt and Craig Joseph, "Intuitive Ethics: How Innately Prepared Intuitions Generate Culturally Variable Virtues," *Daedalus* 133, no. 4 (2004).

unconscious with roots in the evolutionary past of human beings.⁶³ Psychologists are apt to understand human beings in light of dual process models: many cognitive processes occur fast and automatically and outside of conscious awareness; however, human beings can also deliberate about things in a more explicit and conscious manner. The notion of dual processing systems and the fact that many moral judgments seem to be driven by emotion-laden intuitions raises important questions about how reason functions in moral judgment. Jonathan Haidt has developed a social intuitionist account of moral judgment that seeks to account for the complex interplay between reason and the passions in moral judgment. Haidt offers his social intuitionist model as an explicit alternative to rationalist models of moral psychology; however, reason does not disappear in this account. It simply is given a more complex role. Haidt proposes a six-step model of moral judgment. The model is offered here not as a definitive guide to moral judgment but as a helpful heuristic tool for thinking about emotion, reason and their respective roles in moral judgment. Haidt proposes six links in moral judgment:

- (1) Intuitive Judgment Link
- (2) Post-hoc Reasoning Link
- (3) Reasoned Persuasion Link
- (4) Social Persuasion Link
- (5) Reasoned Judgment Link
- (6) Private Reflection Link⁶⁴

The evidence reviewed above concerning the primacy of emotion in moral judgment supports the claim that moral judgments are primarily driven by affect-laden intuitions.

⁶³See Timothy D. Wilson, *Strangers to Ourselves: Discovering the Adaptive Unconscious* (Cambridge, MA: Belknap Press of Harvard University Press, 2002).

⁶⁴Jonathan Haidt, "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment," *Psychological Review* 108, no. 4 (2001).

The primacy of the Intuitive Judgment Link stands in contrast to traditional models of rationalist moral psychology which understand moral judgments to be caused exclusively by reasoned judgment and private reflection. Haidt rejects the rationalist model on the basis of evidence for the primacy of emotion in moral judgment, such as that reviewed above; however, he also rejects this model on the basis of evidence concerning the nature of reasoning. Rationalist models cannot account for the dual-process nature of human cognition, and the post-hoc biased nature that much of our reasoning takes. Haidt's model does not deny a causal role to reason in moral judgment, but offers a more complex account of how that causality occurs. Haidt's model also does not deny the role of reasoned judgment or private reflection, but simply points to their rarity.

Emotion, Post-Hoc Reasoning, and Morality:

Joshua Greene's Critique of Rationalist Deontology

The notion of dual-processing systems and the fact that many moral judgments seem to be driven by affect-laden intuitions raise important questions about how reason functions in moral judgment. One striking discovery in contemporary cognitive science and neuroscience is that when a causal explanation of a behavior is hidden from conscious introspection, human beings will generate a rational explanation for the behavior in question. Evidence for this post-hoc, confabulatory reason comes from both neuroscience and social psychology. Perhaps the most dramatic example of post-hoc confabulatory reason comes from studies of so-called "split brain" patients. These patients had their corpus collusums severed surgically to prevent severe epileptic seizures. Roger Sperry, Joseph Bogen, and Michael Gazzaniga did ground breaking work

on such patients.⁶⁵ Some of the most important findings were discovered in presenting images to the two separated hemispheres of the brain. For instance, an image of a hat presented to subjects' right visual field is processed by their left hemisphere, the hemisphere that also controls the speech centers of the brain. Thus if an image of a hat is flashed to the left hemisphere, when asked to respond, the person would say "hat;" however, if an image was flashed to the right hemisphere—which does not control speech—when asked to respond, the patient—responding from the verbal left hemisphere would say nothing. When asked to point with their left hand to the image, the subject would point to the image of a hat.⁶⁶ Some of the most interesting experimental results occurred when Gazzaniga flashed images simultaneously to the right and left hemispheres. A picture of a chicken claw would be flashed to the left hemisphere while a picture of a snowy scene would be flashed to the right. When asked to point to what they had just seen, the person would point to a chicken with his right hand and to a shovel with his left hand. When the patient was asked to explain his two responses, the patient, responding from the verbal left hemisphere, remarked: "Oh that's easy. The chicken claw goes with the chicken, and you need a shovel to clean out the chicken shed."⁶⁷ Importantly, Gazzaniga argues that the confabulatory/interpretive role of the linguistic centers of the left hemisphere is not unique to split-brain patients. Gazzaniga refers to the

⁶⁵M. Gazzaniga, J. Bogen, and R. Sperry, "Some Functional Effects of Sectioning the Cerebral Commissures in Man," *Proceedings of the National Academy of Sciences, USA* 48 (1962).

⁶⁶Ibid.

⁶⁷Michael Gazzaniga, *The Social Brain* (New York: Basic Books, 1985), 72.

linguistic centers on the left side of the brain as the “interpreter module;” its job is to “interpret” behaviors even though it has no access to the real causes of such behaviors.⁶⁸

A great deal of social psychological research has uncovered a similar pattern of using reason to construct a plausible explanation for a behavior whose actual causal source is hidden from consciousness. In an early experiment of this sort, Richard Nisbett and Timothy Wilson set up a display of five identical pairs of panty hose in a row in a store and asked shoppers to select a pair of the pantyhose. Subjects tended to pick from the right side of the display, but when asked as to why they picked the pair that they did, they offered responses emphasizing the superiority of their selection—knit, sheerness, etc.—however, the actual choice had nothing to do with these factors because the items were in fact identical.⁶⁹

Joshua Greene has recently applied evidence concerning both the emotional nature of moral judgments and the post-hoc confabulatory character of much of our reasoning to philosophical theories of morality. Greene is particularly interested in contrasting deontological and consequentialist patterns of moral judgment. Deontology and consequentialism are both *rationalist* theories of morality. Typically using the language of rights and duties, deontology holds that certain actions are right or wrong irrespective in and of themselves and irrespective of consequences. The criteria that one uses to reach deontological judgments are rational criteria. Consequentialist or teleological accounts of morality assess the moral value of an action solely on the basis of

⁶⁸Ibid.

⁶⁹Richard Nisbett and Timothy Wilson, "Telling More Than We Can Know: Verbal Reports on Mental Processes," *Psychological Review* 84 (1977).

its consequences. Formally articulated forms of deontology and consequentialism have been around for around 200 years; however, an evolutionary perspective views human morality within a much broader scope. In light of this evolutionary perspective, Greene proposes that there may be more to deontology and consequentialism than meets the eye.

...I propose...that the terms “deontology” and “consequentialism” refer to *psychological natural kinds*. I believe that consequentialist and deontological views of philosophy are not so much philosophical inventions as they are philosophical manifestations of two dissociable psychological patterns, two different ways of moral thinking, that have been part of the human repertoire for thousands of years. According to this view the moral philosophies of Kant, Mill, and others are just the explicit tip of large, mostly implicit psychological icebergs. If that is correct, then philosophers may not really know what they’re dealing with when they trade in consequentialist and deontological moral theories, and we may have to do some science to find out.⁷⁰

Greene, in particular, points out that scientific data raises significant questions about the purported rational credentials of deontological moral philosophy. Recall that deontology holds that certain actions are right or wrong irrespective of their consequences, and that reason provides the basis for such moral judgments. Greene presents evidence from a number of sources that show a correlation between deontological judgments and alarm-like emotional responses. Greene proposes that deontological judgments are in fact driven by emotion rather than reason and that deontological reasoning/philosophy serves as a post-hoc justification of an affectively valenced moral intuition. Evidence of correlation between emotion and deontological moral judgments comes from neuroimaging studies of moral dilemmas, studies evaluating attitudes toward punishment, and studies concerning the moral condemnation

⁷⁰Greene, 38.

of harmless actions. Greene's own empirical research centers on the neuroimaging of people engaging various moral dilemmas, such as the "trolley" and "footbridge" problem.

In the trolley dilemma, a runaway trolley is headed for five people who will be killed if it continues on its present course.⁷¹ The only way to save the five people is to hit a switch that will move the trolley to a side track where it will kill one person instead of five. In the footbridge dilemma, a runaway trolley also threatens to kill five people—only in this instance you are standing next to a large stranger on a footbridge, and the only way to save the five is by pushing the stranger onto the tracks.⁷² These two moral dilemmas generate quite different conclusions. The consensus in the trolley dilemma is that it is more acceptable to save five lives at the expense of one; however, the consensus with regard to the footbridge dilemma is that one cannot save five lives at the expense of one. In short, people tend to show a consequentialist response to the trolley dilemma and a deontological response to the footbridge dilemma. Why? This question has vexed philosophers. Some have sought to offer a normative justification for distinguishing between the two cases. Thus the trolley dilemma is an example of the principle of double-effect where someone is harmed as a side effect of an action, whereas footbridge involves using someone as a means. However, responses remain the same in the "loop" version of the trolley dilemma. The trolley is headed toward five people, a side track branches off but in this case loops around and would kill the 5 people if the person was not on the sidetrack. As in the trolley case, people judge that turning the trolley is morally

⁷¹Joshua D. Greene, R. Brian Sommerville, Leigh E. Nystrom, John M. Darley, Jonathan D. Cohen, "An Fmri Investigation of Emotional Engagement in Moral Judgment," *Science* 293 (2001).

⁷²Ibid.

acceptable in order to save five lives, but here, as in the footbridge case, the person is used directly as a means.⁷³

Thus in both dilemmas there is a strong intuition that the moral dilemmas are importantly different and yet people have a hard time explaining exactly what the difference is. “How is it that nearly everyone manages to conclude that it is acceptable to sacrifice one life for five in the trolley dilemma but not in the footbridge dilemma, in spite of the fact that a satisfying justification for distinguishing between these cases is extremely hard to find?”⁷⁴ Greene and his colleagues proposed a purely descriptive approach to these issues. Greene’s group hypothesized that the difference between the dilemmas was to be made on emotional rather than rational grounds. The hypothesis was that “up-close and personal” moral situations like the footbridge case are much more emotionally salient than the impersonal situation found in the trolley case. It is an emotional judgment that drives the judgment of the impermissibility of the act. The basis of drawing this personal/impersonal distinction was largely evolutionary:

The rationale for distinguishing between *personal* and *impersonal* forms of harm is largely evolutionary. “Up close and personal” violence has been around for a very long time, reaching far back into our primate lineage....Given that personal violence is evolutionarily ancient, predating our recently evolved capacities for complex abstract reasoning, it should come as no surprise if we have innate responses to personal violence that are powerful but rather primitive. That is, we might expect humans to have negative emotional responses to certain basic forms of interpersonal violence, where these responses evolved as a means of regulating the behavior of creatures who are capable of intentionally harming one

⁷³Greene, "The Secret Joke of Kant's Soul," 42.

⁷⁴Greene, "An FMRI Investigation of Emotional Engagement in Moral Judgment," 2106.

another, but whose survival depends on cooperation and individual restraint.⁷⁵

This prediction was exactly what was observed. Responding to personal moral dilemmas, as compared with impersonal and nonmoral dilemmas produced increased activity in areas associated with social and emotional processing including the posterior cingulate cortex, the medial prefrontal cortex, the amygdala, and the superior temporal sulcus.⁷⁶ By contrast, both impersonal and non-moral dilemmas produced activity in areas associated with “higher cognition”: the dorsolateral prefrontal cortex and inferior parietal lobe.⁷⁷ Greene also tested reaction times regarding the impermissibility/permissibility of actions in the personal and impersonal moral dilemmas. Personal violations involve a strong emotional inclination not to perform the act. Thus ‘no’ responses regarding the permissibility of a personal moral dilemma should be quick while ‘yes’ responses should take longer because one has to overcome initial emotional resistance.⁷⁸ This is exactly what was observed, with no differences in reaction times in the impersonal or non-moral conditions.⁷⁹ The tendency toward consequentialism or deontology can be explained as a function of emotion: There is a tendency toward consequentialism when the emotional response is low and toward deontology when the emotional response is high.

⁷⁵Greene, "The Secret Joke of Kant's Soul," 43.

⁷⁶Greene, "An Fmri Investigation of Emotional Engagement in Moral Judgment."

⁷⁷Ibid.

⁷⁸Ibid.

⁷⁹Ibid.

This same framework can be applied to other dilemmas. Peter Singer, for example, famously raised the question of why people experience a strict obligation to wade into a pond to save a nearby drowning child but no corresponding obligation to donate to organizations helping sick starving children across the globe. Both situations involve a substantial benefit with negligible cost to the person performing the action. This pattern of response is exactly what one would expect under Greene's hypothesis where characteristically deontological responses are elicited in personal contexts evoking an emotional response.

...[P]eople tend to believe, in a characteristically deontological way, that they are within their moral rights in spending their money on luxuries for themselves, despite the fact that their money could be used to dramatically improve the lives of other people. This is exactly what one would expect if (1) the deontological sense of obligation is driven primarily by emotion, and (2) when it comes to obligations to aid, emotions are only sufficiently engaged when those to whom we might owe something are encountered (or conceived of) in a personal way.⁸⁰

A similar pattern can be observed in attitudes toward punishment.

Consequentialist theories of punishment focus solely on the positive consequences of the punishment. Deontological theories focus on retribution, on giving wrongdoers what they deserve. Experimental evidence has shown that people tend to be retributivists and that this retributivism seems to be emotionally driven, and people punish in proportion to the extent that transgressions make them angry.⁸¹ Greene cites Haidt's studies of the moral

⁸⁰Greene, "The Secret Joke of Kant's Soul," 48.

⁸¹Ibid., 50-55. See J. Baron, R. Gowda, and H. Kunreuther, "Attitudes toward Managing Hazardous Waste: What Should Be Cleaned up and Who Should Pay for It?," *Risk Analysis* 13, no. 2 (1993); D. Kahneman, D. Schkade, and C. Sunstein, "Shared Outrage and Erratic Rewards: The Psychology of Punitive Damages," *Journal of Risk and Uncertainty* 16 (1998); J. Baron and I. Ritov,

condemnation of harmless actions as further evidence of the correlation between emotion and deontology. Many people understood a classic deontological violation (failing to keep one's promise to visit one's mother's grave) as morally wrong, and recall that the bother probe was a better predictor of moral condemnation than the harm probe.⁸²

Greene presents in a number of examples where deontological judgments correlate closely with emotion. Someone who defends rationalist deontology has to explain the correlation and as *rationalists* they cannot say that emotions *cause* the deontological judgment.

Thus, in light of these data, there are a series of coincidences for which various rationalist deontologists must account. For example, according to Judith Jarvis Thomson...and Frances Kamm..., there is a complicated, highly abstract theory of rights that explains why it is okay to sacrifice one life for five in the *trolley* case but not in the *footbridge* case, and it *just so happens*, that we have a strong negative emotional response to the latter case to not to the former. Likewise, according to Colin McGinn and Frances Kamm, there is a theory of duty that explains why we have an obligation to help Singer's drowning child but not comparable obligation to save starving children on the other side of the world, and it *just so happens* that we have strong emotional responses to the former individuals but not the latter. According to Kant and many other legal theorists..., there is a complicated abstract theory of punishment that explains why we ought to punish people regardless of whether there are social benefits to be gained in doing so, and it *just so happens* that we have emotional responses that incline us to do exactly that. The categorical imperative prohibits masturbation because it involves using oneself as a means, and it *just so happens* that the categorical imperative's chief proponent finds masturbation really, really disgusting.⁸³

"Intuitions About Penalties and Compensation in the Context of Tort Law," *Journal of Risk and Uncertainty* 7 (1993).

⁸²Greene, "The Secret Joke of Kant's Soul," 55-58. See Haidt, "Affect, Culture, and Morality, or Is It Wrong to Eat Your Dog?"

⁸³Greene, "The Secret Joke of Kant's Soul," 68.

For Greene, the correlation is best explained by regarding deontological moral philosophy as a post-hoc rationalization of an emotionally driven moral intuition.

Deontology, then, is a kind of moral confabulation. We have strong feelings that tell us in clear and uncertain terms that some things *simply cannot be done* and that other things *simply must be done*. But it is not obvious how to make sense of these feelings, and so we, with the help of some especially creative philosophers, make up a rationally appealing story....Deontology, I believe, is a natural “cognitive” expression of our deepest moral emotions.⁸⁴

The conclusions drawn thus far are the following: much of moral judgment is driven by emotional intuitions and that reason often functions as a post-hoc justification for intuitions.

Emotion and Reason in Jonathan Haidt’s Social Intuitionist

Model of Moral Judgment

The notion that moral judgment is driven by emotional intuitions and that reason often functions in a post-hoc manner does not rule out a causal role for reason to play in moral judgments. The relationship between reason and emotion in the Social Intuitionism is complex, and Haidt’s model allows for reason to play a causal role in several ways.

Recall that Haidt’s model is a six-step approach to moral judgment:

- (1) Intuitive Judgment Link
- (2) Post-hoc Reasoning Link
- (3) Reasoned Persuasion Link
- (4) Social Persuasion Link
- (5) Reasoned Judgment Link
- (6) Private Reflection Link⁸⁵

⁸⁴Ibid., 63.

⁸⁵Haidt, "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment."

We have discussed the first two links in Haidt's model: (1) Intuitive Judgment Link, (2) Post-Hoc Reasoning Link. These links emphasize the *intuitive* aspects of the Social Intuitionist Model. Importantly, Haidt's model does not deny that reason may play an important role in moral judgment. His model includes a role for both (5) Reasoned Judgment and (6) Private Reflection, aspects of moral judgment emphasized in much traditional philosophy and in rationalist moral psychology. While these more private forms of moral reasoning may play a causal role in moral judgment, Haidt argues that these links are utilized relatively infrequently. Haidt argues that reason tends to exercise a causal role primarily within a social context. The role of social aspects of the Social Intuitionist Model are captured in the (3) Reasoned Persuasion Link and (4) Social Persuasion Link. This sensitivity to others and to group norms should not come as a surprise given the ultra social nature of *homo sapiens*.

Only human beings cooperate widely and intensely with nonkin, and we do it in part through a set of social psychological adaptations that make us extremely sensitive to and influenceable by what other people think and feel. We have an intense need to belong and fit in, and our moral judgments are strongly shaped by what others in our 'parish' believe, even when they don't give us any reason for this belief.⁸⁶

Given this context, the moral intuitions of others can influence one's own moral intuitions, even if no explicit reasoning is used. This dynamic is captured in what Haidt calls Social Persuasion. The social intuitionist model also allows reason to play a causal role in moral judgment. He proposes that the reasoning used to justify one's moral judgment can exert a causal effect on the moral intuitions of others. This is the Reasoned

⁸⁶Jonathan Haidt and Fredrik Bjorklund, "Social Intuitionists Answer Six Questions About Moral Psychology," in *Moral Psychology: Volume 2: The Cognitive Science of Morality: Intuition and Diversity*, ed. Walter Sinnott-Armstrong (Cambridge, MA: MIT Press, 2008), 192-193.

Persuasion Link. Haidt argues that because of the strongly affective nature of morality, reasoned persuasion will work more by triggering affectively valenced intuitions in the listener than by providing logically compelling arguments. As noted above, the social intuitionist model also allows for more private forms of reasoning—cases where people reason their way to a conclusion to override an initial intuition. One can see this for example in Greene’s discussion of response latencies in personal, emotion-involving moral dilemmas. However, “such reasoning is hypothesized to be rare, occurring primarily in cases where the initial intuition is weak and the processing capacity is high.”⁸⁷ The Private Reflection Link allows for the spontaneous activation of new intuitions and role taking with regard to these new intuitions. Thus, a person comes to see an issue or dilemma from more than one side and thereby experiences conflicting intuitions. Thus while highlighting the centrality of emotional intuitions in moral judgment, three of the six links in Haidt’s Social Intuitionist Model allow for reasoning to play a causal role: (3) Reasoned Persuasion, (5) Reasoned Judgment, (6) Private Judgment.

The Biological and Cultural Origins of Moral Intuitions

Where do such moral intuitions and the judgments and virtues that they generate come from? An appealing aspect of the social intuitionist model is that it overcomes dichotomies between nativism and empiricism in thinking about the origins of morality. Haidt understands intuitions as having their origin in biology and evolution but also as profoundly shaped by immersion in various culture complexes. As such intuitions serve

⁸⁷Ibid.

as a bridging concept between higher level social science and lower level biology. What sorts of moral intuitions would human beings be prone to develop? Haidt and Joseph surveyed five major contemporary analyses of cross cultural universals and tallied the most frequently mentioned social situations where people were prone to respond either positively or negatively. The winners were suffering/compassion, reciprocity/fairness, hierarchy/respect, and purity/disgust. It is not hard to see the adaptive functions that each of these modules could perform.

Reflecting on suffering, hierarchy, and reciprocity Haidt and Joseph write:

...[T]he prolonged dependence characteristic of primates, especially humans, made it necessary, or at least beneficial, for mothers to detect signs of suffering and distress in their offspring...Psychological preparation for hierarchy evolved to help animals living in social groups make the most of their relative abilities to dominate others....Similarly, a readiness for reciprocity evolved to help animals, particularly primates, reap the benefits of cooperating with non-kin.⁸⁸

Human morality is thus not infinitely malleable but neither is it set in stone.

There are several reasons for this variability. First, drawing upon Dan Sperber's distinction between proper domain and actual domain—the proper domain is the original adaptive trigger; the actual domain is anything that triggers the module—the four moral modules can extend well beyond their original adaptive triggers. For example, disgust, can be extended beyond contaminants and waste products to whole social groups. Secondly, different cultures make relative use of the four modules, and it is not possible for any one culture to emphasize all of the intuitions. Haidt sees the acquisition of morality as similar to the acquisition of phonology. Just as no one language can use all

⁸⁸Haidt, "Intuitive Ethics: How Innately Prepared Intuitions Generate Culturally Variable Virtues," 59.

the clicks, labials, and glottal stops, “a culture that emphasized all of the moral intuitions would risk paralysis as every action triggered multiple conflicting intuitions.” Thus cultures tend to emphasize the different modules to varying degrees

Haidt’s Social Intuitionist Model of moral judgment offers a robust account of moral pluralism, conflict, and diversity to augment the account developed in the earlier sections of this chapter. There we saw that people tend to see divine commands as essential to moral judgment, when in fact they are not, even in very religious communities. Likewise, people tend to see moral properties as part of the objective furniture of the universe, when in fact they are most likely response-dependent properties that depend on certain emotions. The Social Intuitionist Model shows how reason as well as religion can provide a potent source of moral conflict. We think of moral judgments as caused by reason when in fact reasoning may come in after the judgment has been made. As Jonathan Haidt has noted we are prone to two profound illusions with regard to the moral life. First, we tend to believe that our moral reasoning (tail) drives moral judgment (dog) (wag-the-dog illusion). Secondly, we tend to believe that successful rebuttal of arguments will change other’s moral opinions (wag the other dog’s tail illusion).

The bitterness, futility, and self-righteousness of most moral arguments can now be explicated...[In a debate about morality] both sides believe that their positions are based on reasoning about the facts and issues involved (wag-the-dog illusion). Both sides present what they take to be excellent arguments in support of their positions. Both sides expect the other side to be responsive to such reasons (the wag-the-other-dog’s-tail-illusion). When the other side fails to be affected by such good reasons, each side concludes that the other side must be closed minded or insincere. In this way, the culture wars over issues such as homosexuality and

abortion can generate morally motivated players on both sides who believe that their opponents are not morally motivated.⁸⁹

Moral intuitions and virtues are shaped by culture and differ profoundly cross culturally, according to the degree to which the various intuitive building blocks of morality—suffering, reciprocity, hierarchy, purity—are emphasized. Knowledge of these factors alone is not sufficient to resolve issues, but they create an awareness of the complexities involved in moral matters and rule out simplistic solutions.

Mind Science, Metaethics, and Religion III: Implications of Scientific Moral Psychology for Normative Morality and Religious Practices

We have ruled out an approach that sees morality as flowing from religion, specifically from divine commands. Contemporary natural and social science calls into question such claims concerning moral rationalism. Moral judgments are driven by emotion-laden intuitions and profoundly shaped by social and cultural interaction. Where does this leave issues of normative morality and religion? It is argued that the findings of scientific moral psychology support a virtue ethics as the best approach to normative morality. Worries about the normative implications of both the Social Intuitionist Model of moral judgment and virtue theory are then addressed. Finally, it is argued that the focus on emotion and virtue may open the door to a positive valuation of certain religious practices and a fruitful point of contact between religion and science. This claim is investigated in light of Richard Davidson's work on Buddhist mindfulness meditation.

⁸⁹Haidt, "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment," 823.

Scientific Moral Psychology, Virtue Ethics, and
Issues of Normative Moral Evaluation

In terms of the issue of normative implications, a number of naturalistic moral philosophers and scientific moral psychologists have argued that a virtue based understanding of morality is most adequate in light of our scientific understanding of the human person. Jonathan Haidt endorses the turn to virtue theory in several places.

We believe that virtue theories are the most psychologically sound approach to morality....Such theories fit more neatly with what we know about moral development, judgment, and behavior than do theories that focus on moral reasoning or the acceptance of higher level moral principles such as justice.⁹⁰

In another location Haidt writes:

I believe that this turn from character to quandary was a profound mistake for two reasons. First, it weakens morality and limits its scope. Where the ancients saw virtue and character at work in everything a person does, our modern conception confines morality to a set of situations that arise for each person only a few times in a given week” tradeoffs between self-interest and the interests of others. In our thin and restricted modern conception, a moral person is one who gives to charity, helps others, plays by the rules, and in general does not put her own self-interest too far ahead of others....

The second problem with the turn to moral reasoning is that it relies on bad psychology. Many moral education efforts since the 1970’s take the rider off the elephant and train him to solve problems on his own. After being exposed to hours of case studies, classroom discussions about moral dilemmas, and videos about people who faced dilemmas and made the right choices, the child learns how (not what) to think. Then class ends, the rider gets back on the elephant, and nothing changes at recess.⁹¹

The philosopher and moral psychologist William Casebeer makes a similar point.

“Ultimately, the correct situation makes the moral psychology that is required by virtue

⁹⁰Haidt, "Intuitive Ethics: How Innately Prepared Intuitions Generate Culturally Variable Virtues," 62.

⁹¹Haidt, *The Happiness Hypothesis: Finding Modern Truth in Ancient Wisdom*, 164-165.

theory the most neurobiologically plausible, although this is a tentative, defeasible conclusion, and more work is needed to confirm it.”⁹²

The task of normatively evaluating theories grounded in moral emotions and virtue becomes more difficult but not an impossibility. What is lost is simply the false clarity that theological and rationalist approaches aspire to. Theological and rationalist approaches have dominated Western approaches to ethics, and frequently these two approaches are knit together. Both of these approaches involve commitment to strong forms of moral objectivity. One consequence of this is when such views collapse, people tend to think that the only alternative left is subjective relativism. Descriptive accounts of morality, such as Haidt’s social intuitionist model, raise a number of worries for normative morality. In Haidt’s account, virtues are the products of biologically based intuitions shaped by culture. Both the *social* and the *intuitionist* aspects of morality raise concerns. The worry about *intuitionism* is that there is no way to check and transform intuitions and no role for reasoning to play in moral judgment. The worry about *social* is that virtues are culturally specific and that there is no way of evaluating cross cultural claims.

Owen Flanagan has done important work on naturalistic ethics that dovetails with many of the findings discussed above. The worry about the relativity of virtue to social context Flanagan refers to as the “internalist” objection. “The only measure of what constitutes flourishing, excellence, enlightenment is in fact what norms are avowed,

⁹²William D. Casebeer, "Moral Cognition and Its Neural Constituents " *Nature Reviews Neuroscience* 4 (2003).

practiced, considered best from *inside* the culture.”⁹³ However, this state of affairs does not preclude cross-cultural comparison and evaluation; it simply represents honestly the epistemic situation human beings find themselves in. What is precluded is not cross-cultural comparison and evaluation; what is precluded is doing such an analysis from an impossible transcendental standpoint. As Flanagan notes:

We are always because it is the nature of the human condition, in some kind of internalist predicament. But normative assessment, moral improvement, and so on, must come from the dialectical space of our own norms, the norms of those who live differently, and in the space of meta-norms for resolving disagreements or deciding to live tolerantly without agreement. This means that the internalist predicament admits of degrees and that to the degree that we reflect widely, from a perspective that includes other moral conceptions, we are going external in some meaningful sense of the term ‘external’.⁹⁴

In light of this, Flanagan proposes meta-norms for addressing and evaluating moral issues from a vantage point broader than one’s own culture. What Flanagan calls his key meta-norm involves engaging in a process of wide reflective equilibrium (WRE) when a key issue is at stake in terms of norms, virtues, etc.

WRE is a normative test that says we ought to test our ideas about life by bringing them into the widest space of reasons possible. The test, in order to be psychologically realizable, involves taking as genuine all credible contending options available in the space of meaning. Or, perhaps more credible and judicious, it requires reflecting widely when there are misgivings about norms, values, and virtues and when internal scrutiny does not yield a satisfactory solution.⁹⁵

⁹³Owen Flanagan, *The Really Hard Problem: Meaning in a Material World* (Cambridge, MA: MIT Press, 2007), 118.

⁹⁴*Ibid.*, 138.

⁹⁵*Ibid.*, 141.

Another norm Flanagan suggests addresses the worry about the *intuitive* nature of many of our moral judgments. Flanagan calls this an “intuition-checking meta-norm (ICMN):

If nature has gifted us with a moral system that operates mostly intuitively and if intuitions are not always reliable, then we positioned to propose another meta-norm that instructs us to check and double check intuitions and gut reactions: Pay close attention to your intuitive moral responses and to the confidence you experience about the validity of your norms and values. Consider alternatives. Let us call this the intuition checking meta-norm.⁹⁶

This foregoing analysis has called into question many of the metaethical commitments found within religious traditions, but an emphasis on moral emotions and virtue may provide a fruitful point of contact between religion and science. Indeed, much of the Buddhism-mind science dialogue has taken this direction. The final section of this chapter examines this approach, focusing especially on the work of Richard Davidson.

A Fruitful Approach to Religion and Science: The Buddhism-Mind Science Dialogue on Meditation, Affective Style, and Emotion

Contrary to the impression of some Westerners, Buddhism is not a kind of paleo-naturalistic philosophy. At least in its Tibetan versions, it involves what—at least to Western eyes—amount to strong dualist (or idealist) claims. These claims have already been illustrated in the writings of the Dalai Lama and B. Alan Wallace. This dissertation argued that these claims are as problematic as the dualist claims advanced within Western monotheistic traditions. Interestingly, however, there has been a substantial engagement between contemporary mind science and Buddhism where these ontological claims about the nature of mind do not figure prominently at all. Where the dialogue between

⁹⁶Ibid., 142.

Christianity and the mind science has focused almost exclusively on metaphysical issues, the Buddhist-dialogue centers not so much on the metaphysical and doctrinal aspects of religion as its moral and experiential dimensions.⁹⁷ Meditation rather than metaphysics tends to mediate the Buddhism-mind science dialogue.

As was illustrated in the previous chapter, claims about Buddhist meditation are complex, and in many cases there are strong metaphysical claims embedded in the notion of meditation. For instance, Chapter Three analyzed B. Alan Wallace's claims that meditation shows consciousness to be non-physical and that meditation shows that the claims of Tibetan Buddhism are true. However, in much of the Buddhism-mind science dialogue these metaphysical claims are bracketed in favor of a focus on the physical, mental, and moral benefits of mindfulness meditation. This stems from the awareness that there are certain aspects of Buddhist meditation that are amenable to scientific study and certain aspects that are not. Christian dialogue tends to focus on the integration of metaphysical beliefs with the claims of science. Buddhist dialogue focuses on practices and brackets metaphysical issues in the dialogue itself, focusing on issues that are publically accessible and empirically tractable. Here the dialogue sets up a scientific research program to study religious practices. Thus, much of the Buddhism-mind science dialogue itself occurs on naturalistic terms. Chapter Four has already made a case for not just bracketing but for suspicion regarding the metaphysical claims of religious dualism, but here the important point is that scientific and religious parties agree to a methodological bracketing of these claims in creating a scientific research program.

⁹⁷One gets a real sense of these differences when comparing Vatican sponsored religion and science dialogues and dialogues involving the Dalai Lama.

Dialogue takes the form of a scientific research program rather than a discussion of the implications of science of Buddhist metaphysics.

The neuroscientist Richard Davidson has developed a scientific research program for the study of meditation to perhaps the greatest extent. In a discussion of the study of meditation from a scientific perspective, Davidson and his colleagues Antoine Lutz and John Dunne point to the need to tease apart: (1) close descriptions of meditative techniques and states, (2) the metaphysical or soteriological requirements that must be met by these states as expressed in authoritative textual sources.⁹⁸ They use the Tibetan Buddhist practice of “Open Presence” as an example. On the one hand, discussion of this practice includes descriptions of techniques and the phenomenal experiences that should follow. “For example, discursive techniques for de-emphasizing the objectification of sensory content... and the consequent loss of a sense of subject-object duality.”⁹⁹ According to Lutz, Dunne, and Davidson: “these parts of the traditional account lend themselves to investigation inasmuch as they describe techniques and results for which neural correlates may be plausibly postulated and tested.”¹⁰⁰ However, additional metaphysical or ontological claims are often made about such states:

Buddhist philosophical concerns also demand that the state of open presence reflects the ontological foundation of all reality, and Buddhist notions of nirvana also require that the realization of that state will lead the adept to attain inconceivable physical and mental powers....From a

⁹⁸John D. Dunne, Antoine Lutz, and Richard J. Davidson, "Meditation and the Neuroscience of Consciousness: An Introduction," in *The Cambridge Handbook of Consciousness*, ed. P.D. Zelazo, Thompson, E. (Cambridge, UK: Cambridge University Press, 2007), 499.

⁹⁹Ibid.

¹⁰⁰Ibid.

neuroscientific perspective, however, these claims do not lend themselves readily to analysis or description. Thus, from the vantage point of the researcher who stands outside the tradition, it is crucial to separate the highly detailed and verifiable aspects of traditional knowledge about meditation from the transcendental claims that form the metaphysical or theological context of that knowledge.¹⁰¹

Interestingly, this approach is endorsed not only by Western scientists but by figures within the Tibetan Buddhist tradition. Reflecting on a dialogue between Tibetan Buddhists and Western scientists on the topic of “transforming destructive emotions,” the Dalai Lama expresses a willingness to separate the practice of meditation from its embedding within a Buddhist context:

A central claim of Buddhist practice is to reduce the power of destructive emotions in our lives. With that aim in mind, Buddhism offers a wide range of theoretical insights and practical methods. If any of these methods can be shown through scientific tests to be of benefit, then there is every reason to find ways to make them available to everyone, *whether or not they are interested in Buddhism itself*.¹⁰²

Richard Davidson is concerned to take up questions about well-being, happiness, and flourishing in an empirically tractable manner and also to see how certain practices—specifically mindfulness meditation—can lead to changes in the regions of the brain implicated in affective style and moral judgment. An important discovery in recent affective neuroscience and psychology is that an individual’s happiness is constrained in important ways by a happiness set point or their affective style, the “consistent

¹⁰¹Ibid., 500.

¹⁰²Daniel Goleman, *Destructive Emotions: How Can We Overcome Them: A Scientific Dialogue with the Dalai Lama* (New York: Bantam Books, 2004), xiv.

differences in individuals with regard to emotional reactivity and regulation.”¹⁰³

Certainly, the broad notion of affect style is nothing new: Hippocrates and Galen offer interesting discussions of affective personality differences. However, until quite recently, the study of affective style has been based almost exclusively on phenomenological self-reports. This extends to the taxonomy of psychiatric disorders as listed in the DSM-IV.

Much of the work of Davidson’s lab is concerned with penetrating below the phenomenological level.

Many of the parameters of affective style, such as the threshold to respond, magnitude of response, latency to peak response, and recovery function, are features that are often opaque to conscious report, though they may influence the subjective experience of emotion. These parameters of responding can be measured in many different response systems including both the central and peripheral systems....In previous work, we have argued that variations in some of these parameters in particular response systems are especially relevant to vulnerability to mood, anxiety, and other disorders and also to resilience....One of the important developments in emotion research in general, and in affective neuroscience, in particular, is the capacity to objectively measure these parameters of responding.¹⁰⁴

Davidson and his colleagues have been particularly concerned to identify the neural correlates of affective style and to raise the question of the degree to which affective style might be adjusted. Davidson’s group links affective style with the ratio of left-to-right activation in the frontal/anterior regions of the brain, especially the pre-frontal cortex (PFC). In short, greater leftward activation is associated with a more positive affective style while greater rightward activity is associated with a more negative

¹⁰³Richard J. Davidson, "Well-Being and Affective Style: Neural Substrates and Biobehavioral Correlates," *Royal Society London* 359, no. 1449 (2004): 1395.

¹⁰⁴*Ibid.*, 1396.

affective style.¹⁰⁵ Evidence for the relationship between left-to-right PFC activation and affective style comes from lesion studies and EEG and fMRI studies of normal brains. Depressive symptoms tend to be manifested in patients with LPFC damage. The common interpretation of this is that the LPFC is responsible for certain aspects of positive emotion and capacity for positive emotion is impaired with damage to this brain region. Studies with normal subjects have found that LPFC-RPFC can be shifted in lawful ways by inducing negative or positive affect.¹⁰⁶ Further, extreme LPFC and RPFC activity predicts one's dispositional affect as measured in scales like the PANAS and BIS/BAS. PFC activity also predicts one's reaction to certain emotional situations (emotional reactivity). Subjects with strong LPFC activity experienced positive film clips more positively and subjects with strong RPFC activity experienced negative movie clips more negatively.¹⁰⁷

Davidson's group is also interested in the question of the degree to which PFC activity can be altered—its neuroplasticity. Tentative indications of the neuroplasticity of emotional circuitry can already be found in studies of animals and humans. Such research on the plasticity of the brain's emotional circuits is still in its infancy. “Whether repeated practice in techniques of emotion regulation lead to more enduring patterns of brain activation is a question that has not been answered.”¹⁰⁸

¹⁰⁵Ibid.

¹⁰⁶Ibid.

¹⁰⁷Ibid.

¹⁰⁸Ibid.

Davidson and his colleagues tested the effects of an 8-week course on a group of novice meditators from a biotechnology firm in Madison, Wisconsin versus a control group placed on a waiting list.¹⁰⁹ The training consisted of a class that met weekly for two point five hours, home practice, and one seven-hour retreat given during the sixth week of the program. Brain electrical activity was measured by EEG before, immediately after, and four months after the training period. Subjects in the meditative group showed both a significant difference in trait negative affect and showed significantly larger increases in left sided anterior activation.¹¹⁰ There was also a significant difference in immune response in relationship to the leftward shift. Subjects received an influenza vaccine after completing the program. Influenza antibody titers were significantly higher in the meditators than in the control group. Subjects with the largest lefthand shift in the brain in fact showed the largest increase in antibody titers.¹¹¹

In another study, Davidson and his colleagues examined plasticity in some of the neural circuitry associated with empathy and compassion. This study builds on research linking the insula and anterior cingulate gyrus with empathic response. Davidson and colleagues were concerned to investigate the effects of meditation on these areas of the brain.

To cultivate [compassion] practitioners in a number of traditions have developed meditative practices, which are thought to be essential to

¹⁰⁹Richard J. Davidson, Jon Kabat-Zinn, Jessica Schumacher, Melissa Rosenkantz, Daniel Muller, Saki Santorelli, Ferris Urbanowski, Anne Harrington, Katherine Bonus, and John Sheridan, "Alterations in Brain and Immune Function Produced by Mindfulness Meditation," *Psychosomatic Medicine* 65 (2003).

¹¹⁰Ibid.

¹¹¹Ibid.

counteract self-centered tendencies. Techniques included concentration exercises that train attention, behavioral training such as the practice of generosity, cognitive strategies including reflection on the fleeting nature of the self and empathic strategies such as shifting perspectives from self-oriented to other-oriented, or the visualization of the suffering of others. Traditionally, such mental training comprises years of scholastic study and meditative practice. The long-term goal of meditators undergoing such training is to weaken egocentric traits so that altruistic behaviors might arise more frequently and spontaneously.¹¹²

Davidson and his colleagues scanned the brains of novice and expert meditators while meditating and while at rest and when exposed to negative, positive, and neutral sounds. Davidson and his colleagues predicted and found a group by state by valence interaction. Expert meditators while in the state of meditation showed more activation in the insula and anterior cingulate cortices, especially in response to negative sounds than novice meditators. As Davidson summarizes the tentative conclusions:

Because novices and experts differ in many respects other than simply the extent of meditative training (such as culture of origin and first language), longitudinal research that follows individuals over time in response to compassion training will be needed to further substantiate our findings. It will also be essential to assess the impact of such emotional training on behavioral tasks involving altruism, and more generally, emotional reactivity and regulation. The long-term question is to evaluate whether repeated practice in such techniques could result in enduring changes in affective and social style. The fact that large and systematic changes in brain function were observed in response to auditory emotional stimuli presented during the meditative practice of compassion, and the fact that robust differences were observed between experts and novices suggests that the next steps to evaluate the behavioral impact of this training and to longitudinally assess its effects are warranted.¹¹³

While empirical studies of meditation are still in their infancy, they show that

there are empirically tractable ways to explore issues of religion and science, whatever

¹¹²Julie Brefczynski-Lewis Antoine Lutz, Tom Johnstone, and Richard J. Davidson, "Regulation of the Neural Circuitry of Emotion by Compassion Meditation: Effects of Meditative Expertise," *Plosone* 3 (2008).

¹¹³*Ibid.*

the results. An approach to religion and science that focuses on the cultivation of empathy and compassion may prove more significant than approaches focusing on the defense of problematic metaphysical positions like mind-body dualism. A naturalistic perspective in no way rules out a positive valuation of religious practices.

Conclusion

This chapter began with a discussion of the conclusions reached in the first three chapters of this dissertation and their bearing on issues of religion and science. Having focused largely on the impact of the mind sciences on issues of mind-body dualism and its impact on religion, this chapter shifted from metaphysics to morals, specifically metaethical questions concerning the nature and origin of moral judgment. As the previous chapters had used contemporary mind science to shed light on metaphysical issues concerning mind-body dualism and its impact on religion, this chapter examined the implications of the mind sciences for metaethical questions concerning the nature and origin of moral judgment and its impact on religion. A major point of this exercise was to shed light upon the issues of moral pluralism and moral conflict between religious believers and secular naturalists and between religious believers themselves.

Because divine command ethics figures so prominently in discussions of morality and religion, this issue was treated first in light of contemporary work in moral psychology concerning the moral/conventional distinction. Contrary to popular impressions, it was argued that divine commands are not a likely source of moral judgment, as even very religious children reach moral judgments in a manner that is largely autonomous from divine commands. If this is the case, non-religious citizens can

hardly be faulted for explicitly seeking moral wisdom in a manner that is autonomous from religion. At the same time, secular fears regarding divine command ethics may also be misplaced.

Drawing on work on psychopaths and the moral psychology of disgust, it was then argued that emotion likely drives much moral judgment. The emotional origins of morality make moral properties response dependent properties; however, there is initial evidence that people tend to be objectivists about morality. It was proposed that religion and morality may link up over such issues. It was argued that belief in the objectivity of morality, while often seen as essential to morality, may actually be problematic and contribute to the intractability of moral disputes.

The issues of moral pluralism and the nature of moral conflict were further probed in light of Jonathan Haidt's social intuitionist account of moral judgment and Joshua Greene's analysis of deontological moral philosophy as a post-hoc rationalization of emotional intuitions. A prime source of moral conflict lies in that fact that moral judgments are driven by culturally shaped intuitions with reason playing a post-hoc role. However, that causality is usually interpreted backwards. We think reason drives our moral judgments and that we can change the opinions of others simply through rational argument. When this strategy fails, we conclude that the other side is not morally motivated.

Finally, the implications of scientific moral psychology for normative morality and for religion were discussed. Contemporary moral psychology leads to a focus on virtue ethics. This focus on virtue ethics makes the task of normative evaluation much

more challenging. A model for addressing these issues was discussed through the work of Owen Flanagan. In terms of religion, this chapter has primarily concentrated on how science can shed light on the metaethical claims embedded in many religious traditions. Claims that morality is rooted in divine commands or even objectively grounded in a transcendent realm accessed through reason were viewed with suspicion, but this does not rule out a positive valuation of religious practices. Indeed, this may be one of the most fruitful avenues of religion-science dialogue and illustrated in Richard Davidson's research on Buddhist mindfulness practices

CONCLUSION

Questions about human nature, the existence of God and the soul, and the nature and origin of morality have been central questions in both the Western philosophical and theological traditions. In this sense, the questions that this dissertation addresses are quite traditional; however, the answers that the dissertation proposes break strongly with both inherited theology and a great deal of philosophy. The reason for this break comes from the explosion of natural scientific knowledge in the past three centuries and especially the explosion of knowledge concerning the mind-brain in the past hundred years.

Theologians have sought various ways to respond to the challenges of the sciences. Perhaps the most common approach has simply been to ignore the problem or to fail to recognize that there even *is* a problem. For all of its intellectual sophistication, one can read a great deal of theology today and never realize that there was a scientific revolution, a theory of evolution, a neuroscientific revolution, or that the majority of the world's population is not Christian. Sometimes the lack of scientific knowledge is simply due to disciplinary specialization. Other times, the ignorance is principled.

Like other humanists, theologians are apt to make strong *Geisteswissenschaften/naturwissenschaften* distinctions like those discussed in Chapter Two. Others retreat to a faith commitment that insulates itself from critique from "the world." Still others passionate about the pursuit of social justice see questions about the impact of science upon religion as distractions from the task of liberation. But surely an authentic faith

must be intellectually credible and speak to an audience beyond one's particular confessional community. Just as surely, a passion for justice must be informed by the best scientific data and theories concerning society, the human person, and the world. From this standpoint, it is sobering to note that the intellectual beliefs of theologians are at odds with many of their academic colleagues.

Theologians and religious studies scholars interested in 'religion and science' laudably engage the sciences in often sophisticated and profound ways. These theologians *do* have a sense of the challenges the science and naturalistic philosophy pose for religions today, but there is a line that most will not cross: rejecting a commitment to some form of metaphysical dualism (especially in the case of God). In a sense, theologians have stopped short in their analysis of science. They have attempted to address the challenges of the sciences without addressing the challenge of naturalism. The analysis of this dissertation has shown that naturalism is not a dogma that many scientists and philosophers are mired in. The analysis of this dissertation has shown that claims about metaphysical dualism and metaethical objectivism cannot be sustained. Rather than a negative verdict on theology, this dissertation has provided constraints within which a credible theology might operate. In this understanding, the task is not to preserve traditional dualist beliefs at all costs but to develop theology, religion, and spirituality in ways that are intellectually credible.

Aside from calling dualism into question, this dissertation has also been concerned to spell out what explanation in naturalist terms amounts to. Here a crucial concern was to overcome the dichotomy of autonomy versus reduction. This dissertation

called into question versions of the autonomy thesis premised upon metaphysical dualism, the notion that consciousness cannot be explained in bio-psychological terms because consciousness just is not physical. It has also questioned claims about the autonomy of psychology based upon the doctrine of multiple realizability and claims about the autonomy of the humanities based on a principled *Geisteswissenschaften/Naturwissenschaften* distinction. At the same time, it has called into question strong versions of reductionism and eliminativism that fail to take phenomenal consciousness and higher level disciplines seriously.

The dissertation has also sought to spell out fruitful approaches to the religion and science dialogue beyond the conversation stopping approaches so common today. One of its central points in this regard is that religion and science do conflict over issues of metaphysical dualism; however, this conflict need not be a conversation stopper. Conflict does not preclude respectful arguments over issues like dualism, and, in fact, disagreement is essential to move such arguments forward. To this end, this dissertation spells out the kinds of arguments and considerations dualists would have to address in order to defend their position. Despite its skepticism about dualist commitments, this does not mean that there might not be other interesting conversations about religion and science. In Chapter Four, it was proposed that a conversation about morality might take place on different terms than a conversation about metaphysics. This approach was pursued by looking at the implications of scientific moral psychology for metaethical claims about the origins of morality and moral judgment. The point was not to argue a specific stance on a moral issue or to argue that religious believers or naturalists are more

ethical than the other but to challenge the metaethical assumptions that typically frame such debates. This dissertation has barely scratched the surface on issues concerning religion and science, mind and brain, and their mutual relationship, but it has established a naturalistic framework—at once metaphysical, explanatory, and methodological—in which these questions can be further pursued.

BIBLIOGRAPHY

- Astuti, Rita, and Paul Harris. "Understanding Mortality and Life of the Ancestors in Rural Madagascar." *Cognitive Science* 32 (2008): 713-740.
- Atran, Scott. *In Gods We Trust : The Evolutionary Landscape of Religion*. Oxford ; New York: Oxford University Press, 2002.
- Baillargeon, Renee. "The Acquisition of Physical Knowledge in Infancy: A Summary in Eight Lessons." In *Blackwell Handbook of Childhood Cognitive Development*, edited by U.Goswami. Cambridge, MA: Blackwell, 2002.
- Barbour, Ian. *Religion and Science: Historical and Contemporary Issues*. San Francisco, CA: Harper San Francisco, 1997.
- Baron, J., R. Gowda, and H. Kunreuther. "Attitudes toward Managing Hazardous Waste: What Should Be Cleaned up and Who Should Pay for It?" *Risk Analysis* 13, no. 2 (1993): 183-192.
- Baron, J., and I. Ritov. "Intuitions About Penalties and Compensation in the Context of Tort Law." *Journal of Risk and Uncertainty* 7 (1993): 17-33.
- Baron-Cohen, Simon. *Mindblindness: An Essay on Autism and Theory of Mind*. Cambridge, MA: MIT Press, 1995.
- Barrett, Justin. "Is the Spell Really Broken?: Bio-Psychological Explanations of Religion and Theistic Belief." *Theology and Science* 5, no. 1 (2007).
- Barrett, Justin L. *Why Would Anyone Believe in God?* Walnut Creek, CA: AltaMira Press, 2004.
- Bechtel, William. "Mechanism and Phenomenal Experience: The Heuristic Identity Theory." (1993). Retrieved from: <http://mechanism.ucsd.edu/~bill/research/mercier/3rdlecture.pdf> [accessed 12/10/08]
- Bechtel, William, and McCauley, Robert "Heuristic Identity Theory (or Back to the Future): The Mind-Body Problem against the Background of Research Strategies

- in Cognitive Neuroscience." *Proceedings of the 21st Annual Meeting of the Cognitive Science Society* (1999): 67-70.
- Bechtel, William, and Jennifer Mundale. "Multiple Realizability Revisited: Linking Cognitive and Neural States." *Philosophy of Science* 66 (1999): 175-207.
- Bechtel, William, and Robert McCauley. "Explanatory Pluralism and the Heuristic Identity Theory." *Theory Psychology* 11, no. 6 (2001): 736-760.
- Bechtel, William, Pete Mandik, and Jennifer Mundale. "Philosophy Meets the Neurosciences " In *Philosophy and the Neurosciences: A Reader*, edited by Pete Mandik William Bechtel, Jennifer Mundale, Robert Stufflebean. Oxford: Blackwell, 2001.
- Beilby, James K., ed. *Naturalism Defeated?: Essays on Plantinga's Evolutionary Argument against Naturalism*. Ithaca: Cornell University Press, 2002.
- Bering, Jesse M. "Intuitive Conceptions of Dead Agents' Minds: The Natural Foundations of Afterlife Beliefs as Phenomenological Boundary." *Journal of Cognition and Culture* 2, no. 4 (2002): 263-308.
- Bering, Jesse M., and David F. Bjorklund. "The Natural Emergence of Reasoning About the Afterlife as a Developmental Regularity." *Developmental Psychology* 40, no. 2 (2004): 217-233.
- Bering, Jesse M, Carlos Hernandez Blasi, and David F. Bjorklund. "The Development of 'Afterlife' Beliefs in Religiously and Secularly Schooled Children." *British Journal of Developmental Psychology* 23, no. 4 (2005): 587-607.
- Bering, Jesse M., and Becky D. Parker. "Children's Attributions of Intentions to an Invisible Agent." *Developmental Psychology* 42, no. 2 (2006): 253-262.
- Bering, Jesse M. "The Cognitive Psychology of Belief in the Supernatural." *American Scientist* 94, no. 2 (2006): 142-149.
- . "The Folk Psychology of Souls." *Behavioral and Brain Sciences* 29, no. 5 (2006): 453-498.
- Bermudez, Jose Luis. *Philosophy of Psychology: A Contemporary Introduction*. New York and London: Routledge, 2005.
- Bickle, John. *Psychoneural Reduction: The New Wave*. Cambridge, MA: MIT Press, 1998.

- . *The Stanford Encyclopedia of Philosophy*, 2006.
- Blair, R. "A Cognitive Developmental Approach to Morality: Investigating the Psychopath." *Cognition* 57, no. 1 (1995): 1-29.
- . "Brief Report: Morality in the Autistic Child." *Journal of Autism and Developmental Disorders* 26, no. 5 (1996): 571-79.
- Blair, R., C. Sellars, I. Strickland, F. Clark, A. Williams, M. Smith, and L. Jones. "Theory of Mind in the Psychopath." *Journal of Forensic Psychiatry* 7, no. 1 (1996): 15-25.
- Blair, R. "Moral Reasoning and the Child with Psychopathic Tendencies." *Personality and Individual Differences* 22, no. 5 (1997): 731-39.
- Blair, R., L. Jones, F. Clark, and M. Smith. "The Psychopathic Individual: A Lack of Responsiveness to Distress Cues." *Psychophysiology* 34, no. 2 (1997): 192-98.
- Blair, R. "Psychophysiological Responsiveness to the Distress of Others in Children with Autism." *Personality and Individual Differences* 26, no. 3 (1999): 477-85.
- . "Responsiveness to Distress Cues in the Child with Psychopathic Tendencies." *Personality and Individual Differences* 27, no. 1 (1999): 135-45.
- Blair, R. L. Jones, F. Clark, and M. Smith. "Is the Psychopath 'Morally' Insane?" *Personality and Individual Differences* 19 (1997): 741-52.
- Blanke, Olaf. "Stimulating Illusory Own-Body Perceptions." *Nature* 419 (2002): 269-270.
- Blanke, Olaf, Stephanie Ortigue, Theodor Landis, Margitta Seeck. "Stimulating Illusory Own-Body Perceptions." *Nature* 419 (2002): 269-270.
- Blanke, Olaf, Theodor Landis, Laurent Spinelli, and Margitta Seeck. "Out-of-Body Experience and Autoscopy of Neurological Origin." *Brain* 127, no. 2 (2004): 243-258.
- Blanke, Olaf, Christine Mohr, Christoph M. Michel, Alvaro Pascual-Leone, Peter Brugger, Margitta Seeck, Theodor Landis, Gregor Thut. "Linking out-of-Body Experience and Self Processing to Mental Own-Body Imagery at the Temporoparietal Junction." *Journal of Neuroscience* 25, no. 3 (2005).
- Blanke, Olaf, and Shahar Arzy. "The out-of-Body Experience: Disturbed Self-Processing at the Temporo-Parietal Junction." *Neuroscientist* 11, no. 16 (2005): 16-23.

- . "The out-of-Body Experience: Disturbed Self-Processing at the Temporo-Parietal Junction." *Neuroscientist* 11, no. 1 (2005): 16-24.
- Block, Ned and Robert Stalnaker. "Conceptual Analysis, Dualism, and the Explanatory Gap." *The Philosophical Review* 108, no. 1 (1999): 1-46.
- Bloom, Paul. *Descartes' Baby : How the Science of Child Development Explains What Makes Us Human*. New York: Basic Books, 2004.
- . "Is God an Accident?" *The Atlantic Monthly* (2005).
- . "Religion Is Natural." *Developmental Science* 10, no. 1 (2007): 147-151.
- Botvinick, Matthew, and Jonathan Cohen. "Rubber Hands 'Feel' Touch That Eyes See." *Nature* 391 (1998): 756.
- Boyer, Pascal. *Religion Explained: The Evolutionary Origins of Religious Thought*. New York: Basic Books, 2001.
- Brown, Warren S., Nancey Murphy, and H. Newton Malony, ed. *Whatever Happened to the Soul: Scientific and Theological Portraits of Human Nature*. Minneapolis, MN: Fortress Press, 1998.
- Byrne, Peter. *The Moral Interpretation of Religion*. Grand Rapids, MI: Eerdmans, 1998.
- Brown, Warren S., Nancy Murphy, and H. Newton Malony, ed. *Whatever Happened to the Soul: Scientific and Theological Portraits of Human Nature*. Minneapolis, MN: Fortress Press, 1998.
- Callebaut, Werner. *Taking the Naturalistic Turn or How Real Philosophy of Science Is Done*. Chicago, IL: University of Chicago Press, 1993.
- Casebeer, William D. "Moral Cognition and Its Neural Constituents." *Nature Reviews Neuroscience* 4 (October 2003): 841-846.
- Chalmers, David J. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press, 1996.
- Churchland, Paul. "Chimerical Colors: Some Phenomenological Predictions from Cognitive Neuroscience." *Philosophical Psychology* 18, No. 5 (2005). In *Neurophilosophy at Work*. Cambridge, MA: MIT Press, 2007.
- Churchland, Paul M. *Matter and Consciousness*. Cambridge, MA: MIT Press, 1984.

- Churchland, Patricia Smith. *Neurophilosophy: Toward a Unified Science of the Mind/Brain*. Cambridge, MA: MIT Press, 1986.
- . "The Hornswoggle Problem." (1996). Retrieved from: <http://www.wm-johnston.co.uk/philosophy/hornswoggle.htm> [accessed 11/21/08].
- . *Brain-Wise: Studies in Neurophilosophy*. Cambridge, MA: MIT Press, 2002.
- Clayton, Philip. "Neuroscience, the Person, and God: An Emergentist Account." In *Neuroscience and the Person: Scientific Perspectives on Divine Action*, edited by Nancy Murphy Robert John Russell, Theo C. Meyering, Michael A. Arbib, 181-214. Vatican City State: Vatican Observatory Publications, 2002.
- . *Mind and Emergence: From Quantum to Consciousness*. Oxford: Oxford University Press, 2004.
- Cooper, John. *Body, Soul, and Life Everlasting: Biblical Anthropology and the Monism-Dualism Debate*. Grand Rapids, MI: Eerdmans, 1989.
- Craver, Carl. "Role Functions, Mechanisms, and Hierarchy." *Philosophy of Science* 68, no. 1 (2001): 53-74.
- Craver, Carl F. *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford: Oxford University Press, 2007.
- Crick, Francis. *The Astonishing Hypothesis: The Scientific Search for the Soul*. New York: Simon & Schuster, 1994.
- Csibra, G., G. Gergely, S. Biro, O. Koos, and M. Brockbank. "Goal Attribution without Agency Cues: The Perception of "Pure Reason" In Infancy." *Cognition* 72, no. 3 (1999): 237-267.
- Cunningham, Lawrence S., and John Kelsay. *The Sacred Quest: An Invitation to the Study of Religion*. Upper Saddle River, NJ: Pearson/Prentice Hall, 2006.
- Dalai Lama. *The Universe in a Single Atom: The Convergence of Science and Spirituality*. New York: Morgan Road Books, 2005.
- Dalai Lama and Daniel Goleman. "On the Luminosity of Being." *New Scientist* 178, no. 2396 (2003): 42.
- Damasio, Antonio R. *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Penguin, 1994.

- . *The Feeling of What Happens: Body and Emotion in the Making of Consciousness* (1st ed.). New York: Harcourt Brace, 1999.
- Davidson, Richard, and Antoine Lutz. "Buddha's Brain: Neuroplasticity and Meditation." *IEEE Signal Processing Magazine* (2007): 171-174.
- Davidson, Richard J., Jon Kabat-Zinn, Jessica Schumacher, Melissa Rosenkantz, Daniel Muller, Saki Santorelli, Ferris Urbanowski, Anne Harrington, Katherine Bonus, and John Sheridan. "Alterations in Brain and Immune Function Produced by Mindfulness Meditation." *Psychosomatic Medicine* 65 (2003): 564-570.
- . "Alterations in Brain and Immune Function Produced by Mindfulness Meditation." *Psychosomatic Medicine* 65 (2003): 564-570.
- Davidson, Richard J. "Well-Being and Affective Style: Neural Substrates and Biobehavioral Correlates." *Royal Society London* 359, no. 1449 (2004): 1395-1411.
- Dennett, Daniel C. *Consciousness Explained*. Boston, MA: Little, Brown, 1991.
- Deutsch, Max. "Subjective Physical Facts." Retrieved from: <http://neologic.net/rd/chalmers/mdeutsch.html> [accessed 1/3/09].
- Durkheim, Emile. *The Rules of the Sociological Method*. Glencoe, IL: Free Press, 1895/1962.
- . *The Elementary Forms of Religious Life*. Translated by Joseph Swain. New York: Free Press, 1915/1966.
- Edgell, Penny, Joseph Gertais, and Douglas Hartmann. "Atheists As "Other": Moral Boundaries and Cultural Membership in American Society." *American Sociological Review* 21, no. 2 (2006).
- Edwards, Paul. *Reincarnation: A Critical Examination*. New York: Prometheus Books, 1996.
- Eliade, Mircea. *Patterns in Comparative Religion*. Translated by Rosemary Sheed. Cleveland, OH and New York: World, 1968.
- Evans, E.M. "Beyond Scopes: Why Creationism Is Here to Stay." In *Imagining the Impossible: The Development of Magical, Scientific, and Religious Thinking in Contemporary Society*, edited by C.N. Johnson K.S. Rosengren, and P.L. Harris, 305-333. Cambridge, UK: Cambridge University Press, 2000.

- . "The Emergence of Beliefs About the Origin of Species in School-Age Children." *Merrill Palmer Quarterly* 46 (2000): 221-254.
- . "Cognitive and Contextual Factors in the Emergence of Diverse Belief Systems: Creation Versus Evolution." *Cognitive Psychology* 42, no. 3 (2001): 217-266.
- Farber, Ilya. "How a Neural Correlate Can Function as an Explanation of Consciousness." *Journal of Consciousness Studies* 12, no. 4-5 (2005).
- Feinberg, Todd. *Altered Egos: How the Brain Creates the Self*. Oxford: Oxford University Press, 2001.
- Flanagan, Owen. *The Science of the Mind*. Cambridge, MA: MIT Press, 1991.
- Flanagan, Owen. *Consciousness Reconsidered*. Cambridge, MA: MIT Press, 1992.
- Flanagan, Owen. *The Problem of the Soul: Two Versions of Mind and How to Reconcile Them*. New York: Basic Books, 2002.
- . *The Really Hard Problem: Meaning in a Material World*. Cambridge, MA: MIT Press, 2007.
- Fodor, Jerry. "Special Sciences." In *The Philosophy of Science*, edited by Philip Gasper, Richard Boyd, and J.D. Trout. Cambridge, MA: MIT Press, 1991.
- Gazzaniga, M., J. Bogen, and R. Sperry. "Some Functional Effects of Sectioning the Cerebral Commissures in Man." *Proceedings of the National Academy of Sciences, USA* 48 (1962): 1765-1769.
- Gazzaniga, Michael. *The Social Brain*. New York: Basic Books, 1985.
- Geertz, Clifford, ed. *The Interpretation of Cultures: Selected Essays*. New York: Basic Books, 1973.
- Gergely, G., Z. Nadasdy, G. Scibra, and S. Biro. "Taking the Intentional Stance at 12 Months of Age." *Cognition* 56 (1995): 165-193.
- Goetz, Charles Taliaferro and Stewart. *Naturalism*. Grand Rapids, MI: Eerdmans, 2008.
- Goldman, Alvin I. *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading Philosophy of Mind*. Oxford; New York: Oxford University Press, 2006.

- Goleman, Daniel. *Destructive Emotions: How Can We Overcome Them: A Scientific Dialogue with the Dalai Lama*. New York: Bantam Books, 2004.
- Gopnik, Alison. "Mindblindness." Unpublished Manuscript, Department of Psychology, University of California, Berkeley (1993).
- Gould, Stephen Jay. *Rocks of Ages: Science and Religion in the Fullness of Life*. New York: Ballantine Books, 1999.
- Greene, Joshua, and Jonathan Haidt. "How (and Where) Does Moral Judgment Work?" *Trends in Cognitive Sciences* 6, no. 12 (2002): 517-522.
- Greene, Joshua. "From Neural 'Is' to Moral 'Ought': What Are the Moral Implications of Neuroscientific Moral Psychology?" *Nature Reviews Neuroscience* 4 (October 2003): 847-850.
- Greene, Joshua D., R. Brian Sommerville, Leigh E. Nystrom, John M. Darley, Jonathan D. Cohen. "An Fmri Investigation of Emotional Engagement in Moral Judgment." *Science* 293 (2000): 2105-2108.
- Greene, Joshua D. "The Secret Joke of Kant's Soul." In *Moral Psychology: Volume 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, edited by Walter Sinnott-Armstrong, pp. 35-79. Cambridge, MA: MIT Press, 2008.
- Guthrie, Stewart. *Faces in the Clouds: A New Theory of Religion*. Oxford: Oxford University Press, 1993.
- Haidt, Jonathan, Silvia Helena Koller, and Maria G. Dias. "Affect, Culture, and Morality, or Is It Wrong to Eat Your Dog?" *Journal of Personality and Social Psychology* 65, no. 4 (1993): 613-628.
- Haidt, Jonathan. "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment." *Psychological Review* 108, no. 4 (2001): 814-834.
- Haidt, Jonathan, and Matthew A. Hersh. "Sexual Morality: The Cultures and Emotions of Conservatives and Liberals." *Journal of Applied Social Psychology* 31, no. 1 (2001): 191-221.
- Haidt, Jonathan, and Craig Joseph. "Intuitive Ethics: How Innately Prepared Intuitions Generate Culturally Variable Virtues." *Daedalus* 133, no. 4 (2004): 55-65.
- Haidt, Jonathan. *The Happiness Hypothesis: Finding Modern Truth in Ancient Wisdom*. New York: Basic Books, 2006.

- Haidt, Jonathan, and Fredrick Bjorklund. "Social Intuitionists Answer Six Questions About Moral Psychology." In *Moral Psychology: Volume 2: The Cognitive Science of Morality: Intuition and Diversity*, edited by Walter Sinnott-Armstrong, 181-217. Cambridge, MA: MIT Press, 2008.
- Harris, Paul L., and Marta Gimenez. "Children's Acceptance of Conflicting Testimony: The Case of Death." *Journal of Cognition and Culture* 5, no. 1-2 (2005): 143-164.
- Hauser, Marc. *Moral Minds: How the Nature Designed Our Universal Sense of Right and Wrong*. New York: Harper Collins, 2006.
- Heider, F., and S. Simmel. "An Experimental Study of Apparent Behavior." *American Journal of Psychology* 57, no. 2 (1944).
- Heil, John. *Philosophy of Mind: A Contemporary Introduction*. New York and London: Routledge, 2004.
- Hick, John. *The New Frontier of Science and Religion: Religious Experience, Neuroscience, and the Transcendent*. New York: Palgrave MacMillan, 2006.
- Jackson, Frank. "Epiphenomenal Qualia." In *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*, edited by Yujin Nagasawa Peter Ludlow, and Daniel Stoljar. Cambridge, MA: MIT Press, 2004.
- . "What Mary Didn't Know." In *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*, edited by Yujin Nagasawa Peter Ludlow, and Daniel Stoljar. Cambridge, MA: MIT Press, 2004.
- Johnson, Jesse M. Bering, and Dominic D.P. "'O Lord... You Perceive My Thoughts from Afar": Recursiveness and the Evolution of Supernatural Agency." *Journal of Cognition and Culture* 5, no. 1-2 (2005): 118-142.
- Johnson, S., V. Slaughter, and S. Carey. "Whose Gaze Will Infants Follow?" *Developmental Science* (1998): 233-238.
- Justin Barrett, R. Richert, and A. Driesenga. "God's Beliefs Versus Mother's: The Development of Non-Human Agent Concepts." *Child Development* 72, no. 1 (2001): 50-65.
- Kahneman, D., D. Schkade, and C. Sunstein. "Shared Outrage and Erratic Rewards: The Psychology of Punitive Damages." *Journal of Risk and Uncertainty* 16 (1998).

- Katz, Steven, ed. *Mysticism and Philosophical Analysis*. New York: Oxford University Press, 1978.
- . ed. *Mysticism and Religious Traditions*. New York: Oxford University Press, 1983.
- Keil, Frank. *Concepts, Kinds, and Cognitive Development*. Cambridge, MA: MIT Press, 1989.
- Kelemen, Deborah. "Beliefs About Purpose: On the Origins of Teleological Thought." In *The Descent of Mind: Psychological Perspectives on Hominid Evolution*, edited by M. Corballis and S. Lea, 278-294. Oxford: Oxford University Press, 1999.
- . "The Scope of Teleological Thinking in Preschool Children." *Cognition* 70, no. 3 (1999): 241-272.
- . "Why Are Rocks Pointy? Children's Preference for Teleological Explanations of the Natural World." *Developmental Psychology* 35, no. 6 (1999): 1440-1453.
- . "Why Things Happen: Teleological Explanation in Parent-Child Conversations." In press (2002).
- . "British and American Children's Preferences for Teleo-Functional Explanations of the Natural World." *Cognition* 88, no. 2 (2003): 201-221.
- . "Are Children 'Intuitive Theists'? Reasoning About Purpose and Design in Nature." *Psychological Science* 15, no. 5 (2004): 295-301.
- Kelemen, Deborah, and C. DiYanni. "Intuitions About Origins: Purpose and Intelligent Design in Children's Reasoning About Nature." *Journal of Cognition and Development* (In press).
- Kellman, J., and E.S. Spelke. "Perception of Partly Occluded Objects in Infancy." *Cognitive Psychology* 15, no. 4 (1983): 483-524.
- Kripke, Saul. *Naming and Necessity*. Cambridge, MA: Harvard University Press, 1972.
- . "Identity and Necessity." In *Philosophy of Mind: A Guide and Anthology*, edited by John Heil, pp. 128-136. Oxford: Oxford University Press, 2004.
- Kuhlmeier, V., K. Wynn, and P. Bloom. "Attribution of Dispositional States by 12-Month Olds." *Psychological Science* 14, no. 5 (2003): 402-408.

- Kuhlmeier, Valerie A., Paul Bloom, and Karen Wynn. "Do 5-Month-Old Infants See Humans as Material Objects?" *Cognition* (2004).
- Lenggenhager, Bigna, Tej Tadi, Thomas Metzinger, and Olaf Blanke. "Video Ergo Sum: Manipulating Bodily Self Consciousness." *Science* 317, no. 5841 (2007): 1096-1099.
- Lenggenhager, Bigna, Tej Tadi, Thomas Metzinger, and Olaf Blanke. "Video Ergo Sum: Manipulating Bodily Self Consciousness." *Science* 317, no. 5841 (2007): 1096-1099.
- Leslie, A. "The Theory of Mind Impairment in Autism." In *Natural Theories of Mind*, edited by A. Whiten. Oxford: Blackwell, 1991.
- Levine, Joseph. "Materialism and Qualia: The Explanatory Gap." *Pacific Philosophical Quarterly* (1983): 354-361.
- Llinas, Rodolfo. *I of the Vortex: From Neurons to Self*. Cambridge, MA: MIT Press, 2001.
- Loar, Brian. "Phenomenal States (Revised Version)." In *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument* edited by Yujin Nagasawa Peter Ludlow, and Daniel Stoljar. Cambridge, MA: MIT Press, 2004.
- Lopez, Donald. *Buddhism and Science: A Guide for the Perplexed*. Chicago, IL: University of Chicago Press, 2008.
- Lutz, Antoine, John D. Dunne, and Richard J. Davidson. "Meditation and the Neuroscience of Consciousness: An Introduction." In *The Cambridge Handbook of Consciousness*, edited by P.D. Zelazo, E.Thompson. Cambridge, UK: Cambridge University Press, 2007.
- Lutz, Antoine, Julie Brefczynski-Lewis, Tom Johnstone, and Richard J. Davidson. "Regulation of the Neural Circuitry of Emotion by Compassion Meditation: Effects of Meditative Expertise." *Plosone* 3 (2008).
- Machamer, Peter, Lindley Darden, and Carl Craver. "Thinking About Mechanisms." *Philosophy of Science* 67 (March 2000).
- Mandik, Pete. "Points of View from the Brain's Eye View: Subjectivity and Neural Representation." In *Philosophy and the Neurosciences: A Reader*, edited by Pete Mandik William Bechtel, Jennifer Mundale, Robert Stufflebean. Oxford: Blackwell, 2001.

- McCauley, Robert. "Overcoming Barriers to a Cognitive Psychology of Religion." *Method and Theory in the Study of Religion* 12, no. 1 (2000): 141-161.
- . "Explanatory Pluralism and the Co-Evolution of Theories in Science." In *Philosophy and the Neurosciences: A Reader*, edited by Pete Mandik William Bechtel, Jennifer Mundale, Robert Stufflebean. Oxford: Blackwell, 2001.
- McGinn, Colin. *The Mysterious Flame: Conscious Minds in a Material World*. New York: Basic Books, 1999.
- Meltzoff, A.N., and M.K. Moore. "Imitations of Facial and Manual Gestures by Human Neonates." *Science* 198 (1977): 702-709.
- Metzinger, Thomas. *Being No One: The Self-Model Theory of Subjectivity*. Cambridge, MA: MIT Press, 2003.
- . "Out-of-Body Experiences as the Origin of the Concept of a "Soul"." *Mind and Matter* 3, no. 1 (2005): 57-84.
- . *The Ego Tunnel: The Science of the Mind and the Myth of the Self*. New York: Basic Books, 2009.
- Mohr, Christine, and Olaf Blanke. "The Demystification of Autoscopical Phenomena: Experimental Propositions." *Current Psychiatry Reports* 7, no. 7 (2005): 189-195.
- Moreland, J.P. *Consciousness and the Existence of God*. London: Routledge, 2008.
- Murphy, Nancy. *Bodies and Souls, or Spirited Bodies?* Cambridge, MA: Cambridge University Press, 2006.
- Nagel, Ernest. *The Structure of Science*. New York: Harcourt, Brace and World, 1961.
- Nagel, Thomas. "What Is It Like to Be a Bat?" In *Philosophy of Mind: A Guide and Anthology*, edited by John Heil. Oxford: Oxford University Press, 2004.
- Nansook Park, Christopher Peterson, and Martin Seligman. "Strengths of Character and Well-Being." *Journal of Social and Clinical Psychology* 23, no. 5 (2004): 603-619.
- Newberg, Andrew, and Eugene D'Aquili. *Why God Won't Go Away: Brain Science and the Biology of Belief*. New York: Ballantine Books, 2001.
- Nichols, Shaun. "Norms with Feeling: Towards a Psychological Account of Moral Judgment." *Cognition* 84 (2002): 221-226.

- Nichols, Shaun, and Trisha Folds-Bennett. "Are Children Moral Objectivists? Children's Judgments About Moral and Response-Dependent Properties." *Cognition* 90, no. 2 (2003): B23-B32.
- Nichols, Shaun. "After Objectivity: An Empirical Account of Moral Judgment." *Philosophical Psychology* 17 (2004).
- . *Sentimental Rules: On the Natural Foundations of Moral Judgment*. Oxford: Oxford University Press, 2004.
- Nisbett, Richard, and Timothy Wilson. "Telling More Than We Can Know: Verbal Reports on Mental Processes." *Psychological Review* 84 (1977): 231-259.
- Norenzayan, A., and S. Atran. "Cognitive and Emotional Processes in the Cultural Transmission of Natural and Nonnatural Beliefs." In *The Psychological Foundations of Culture*, edited by M. Schaller and C. Crandall. Hillsdale, NJ: Erlbaum, 2002.
- Nucci, Larry, and Elliot Turiel. "God's Word, Religious Rules, and Their Relation to Christian and Jewish Children's Concepts of Morality." *Child Development* 64, no. 5 (1993): 1475-1491.
- Nucci, Larry P. "Children's Conceptions of Morality, Societal Convention, and Religious Prescription." In *Moral Dilemmas: Philosophical and Psychological Issues in the Development of Moral Reasoning*, edited by Carol Gibb Harding, pp. 137-173. Chicago, IL: Precedent Publishing, 1985.
- Papineau, David. *Philosophical Naturalism*. Oxford: Blackwell, 1993.
- . *Thinking About Consciousness*. Oxford: Oxford University Press, 2002.
- Parfit, Derek. *Reasons and Persons*. Oxford: Oxford University Press, 1984.
- Pinker, Steven. *The Blank Slate: The Modern Denial of Human Nature*. New York: Viking, 2002.
- Plantinga, Alvin. "Advice to Christian Philosophers." *Faith and Philosophy: Journal of the Society of Christian Philosophers* 1 (October 1984).
- . *Warrant and Proper Function*. Oxford: Oxford University Press, 1993.
- . "On Christian Scholarship." (1994). Retrieved from: http://www.calvin.edu/academic/philosophy/virtual_library/articles/plantinga_alvin/on_christian_scholarship.pdf [accessed 7/31/09].

- . "Materialism and Christian Belief." In *Persons: Human and Divine*, edited by Peter van Inwagen and Dean Zimmerman, pp. 99-141. Oxford: Oxford University Press, 2007.
- Polger, Thomas. *Natural Minds*. Cambridge, MA: MIT Press, 2004.
- Pope John Paul II. "Truth Cannot Contradict Truth." (1996). Retrieved from: http://www.newadvent.org/library/docs_jp02tc.htm [accessed 10/10/09].
- Premack, D., and A. Premack. "Origins of Social Competence." In *The Cognitive Neurosciences*, edited by Michael Gazzaniga. Cambridge, MA: MIT Press, 1995.
- . "Infants Assign Value to the Goal-Directed Actions of Self-Propelled Objects." *Journal of Cognitive Neuroscience* 9 (1997): 848-856.
- Prinz, Jesse. *The Emotional Construction of Morals*. Oxford: Oxford University Press, 2007.
- Proudfoot, Wayne. *Religious Experience*. Berkeley, CA: University of California Press, 1985.
- Putnam, Hilary. "The Nature of Mental States." In *Mind and Cognition: A Reader*, edited by William Lycan. Oxford: Blackwell, 1990.
- Pyysiainen, Ilkka, and Marc Hauser. "The Origins of Religion: Evolved Adaptation or by-Product?" *Trends in Cognitive Sciences* 14, no. 3 (2009): 104-108.
- Ramachandran, V.S., and S. Blakeslee. *Phantoms in the Brain*. New York: William Morrow, 1998.
- Ramsey, William. "Naturalism Defended." In *Naturalism Defeated?: Essays on Plantinga's Evolutionary Argument against Naturalism*, edited by James K. Beilby. Ithaca, NY: Cornell University Press, 2002.
- Repacholi, B.M., and A. Gopnik. "Early Reasoning About Desires: Evidence from 14- and 18-Month Olds." *Developmental Psychology* 33 (1997): 12-21.
- Revonsuo, Annti. "Consciousness, Dreams, and Virtual Realities." *Philosophical Psychology* 8 (1995).
- Revonsuo, Antti. *Inner Presence: Consciousness as a Biological Phenomenon*. Cambridge, MA: MIT Press, 2006.

- Robbins, Philip, and Anthony I. Jack. "The Phenomenal Stance." *Philosophical Studies* 127 (2006): 59-85.
- Robert John Russell, Nancy Murphy, Theo C. Meyering, Michael A. Arbib, ed. *Neuroscience and the Person: Scientific Perspectives on Divine Action*. Vatican City State: Vatican Observatory Publications, 2002.
- Schnall, S., J. Haidt, and G. Clore. "Irrelevant Disgust Makes Moral Judgment More Severe, for Those Who Listen to Their Bodies." University of Virginia, 2004.
- Schouten, Maurice K.D. "Theism, Dualism, and the Scientific Image of Humanity." *Zygon* 36, no. 4 (1991).
- Searle, John. *The Rediscovery of the Mind*. Cambridge, MA: MIT Press, 1992.
- Seligman, Martin, and Christopher Peterson. *Character Strengths and Virtues: A Handbook and Classification*. Oxford: Oxford University Press, 2004.
- Sessions, William Lad. *Reading Hume's Dialogues: A Veneration for True Religion*. Bloomington & Indianapolis, IN: Indiana University Press, 2002.
- Sharf, Robert. "Buddhist Modernism and the Rhetoric of Meditative Experience." *Numen* 42 (1995).
- . "Experience." In *Critical Terms for Religious Studies*, edited by Mark C. Taylor. Chicago, IL: University of Chicago Press, 1998.
- Shermer, Michael. *How We Believe: Science, Skepticism, and the Search for God*. New York: Henry Holt, 2000.
- Silberstein, Michael. "Emergence, Theology, and the Manifest Image." In *The Oxford Handbook of Religion and Science* edited by Philip Clayton. Oxford: Oxford University Press, 2006.
- Sinnott-Armstrong, Walter, ed. *Moral Psychology: Volume 2: The Cognitive Science of Morality: Intuition and Diversity*, Vol. 2. Cambridge, MA: MIT Press, 2008.
- . ed. *Moral Psychology: Volume 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development*. Cambridge, MA: MIT Press, 2008.
- . *Morality without God?* Oxford: Oxford University Press, 2009.
- Slingerland, Edward. *What Science Teaches the Humanities: The Integration of Body and Culture*. Cambridge, UK: Cambridge University Press, 2008.

- Spelke, E. "Initial Knowledge: Six Suggestions." *Cognition* 50, no. 1-3 (1994): 443-447.
- Swinburne, Richard. *The Coherence of Theism*. Oxford: Clarendon, 1977.
- . *Is There a God?* Oxford: Oxford University Press, 1996.
- Taliaferro, Charles. *Consciousness and the Mind of God*. Cambridge, UK: Cambridge University Press, 1994.
- Taylor, M. *Imaginary Companions and the Children Who Create Them*. New York: Oxford University Press, 1999.
- Tomasello, M., R. Strosberg, and N. Akhtar. "Eighteen-Month-Old Children Learn Words in Non-Ostensive Contexts." *Journal of Child Language* 23, no. 1 (1996): 157-176.
- Tracy, David. *Blessed Rage for Order: The New Pluralism in Theology*. Chicago, IL: University of Chicago Press, 1976.
- Tremlin, Todd. *Minds and Gods: The Cognitive Foundations of Religion*. New York: Oxford University Press, 2006.
- Turiel, E., M. Killen, and C. Helwig. "Morality: Its Structure, Functions, and Vagaries." In *The Emergence of Morality in Young Children*, edited by J. Kagan and S. Lamb, pp. 155-244. Chicago, IL: University of Chicago Press, 1987.
- Van Gulick, Robert. "Conceiving Beyond Our Means: The Limits of Thought Experiments." *Toward a Science of Consciousness* 3 (2003). Retrieved from: http://cognet.mit.edu/posters/TUCSON3/Van_Gulick.html [accessed 11/17/08].
- . "So Many Ways of Saying No to Mary." In *There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*, edited by Yujin Nagasawa Peter Ludlow, and Daniel Stoljar. Cambridge, MA: MIT Press, 2003.
- van Inwagen, Peter, and Dean Zimmerman, ed. *Persons: Human and Divine*. Oxford: Oxford University Press, 2008.
- Wallace, B. Alan. *The Taboo of Subjectivity: Toward a New Science of Consciousness*. Oxford: Oxford University Press, 2000.
- . ed. *Buddhism and Science: Breaking New Ground*. New York: Columbia University Press, 2003.

- . *Contemplative Science: Where Buddhism and Neuroscience Converge*. New York: Columbia University Press, 2007.
- . *Mind in the Balance: Meditation in Science, Buddhism, and Christianity*. New York: Columbia University Press, 2009.
- Wegner, Daniel M. *The Illusion of Conscious Will*. Cambridge, MA: MIT Press, 2002.
- Wheatley, Thalia, and Jonathan Haidt. "Hypnotic Disgust Makes Moral Judgment More Severe." *Psychological Science* 16, no. 10 (2005): 780-784.
- Wilson, Timothy D. *Strangers to Ourselves: Discovering the Adaptive Unconscious*. Cambridge, MA: Belknap Press of Harvard University Press, 2002.

VITA

A native of Indiana, Pennsylvania, Paul Voelker earned Bachelor of Arts and Master of Divinity degrees from the University of Notre Dame and a Doctor of Philosophy degree from Loyola University Chicago, where this dissertation was defended with an award of distinction. His academic work is at the intersection of religious studies, philosophy, and mind science. In particular, he is interested in approaching the study of religion through the lens of contemporary neuroscience, cognitive science, and philosophy of mind and in developing naturalistic approaches to the philosophy of religion and the religion and science dialogue.