EMERGENCE AND THE PROBLEM OF PERSONAL IDENTITY

A Thesis

Presented to the

Faculty of

San Diego State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

in

Philosophy

by

David Thomas Jerome Neuburger

Spring 2013

SAN DIEGO STATE UNIVERSITY

The Undersigned Faculty Committee Approves the

Thesis of David Thomas Jerome Neuburger:

Emergence and the Problem of Personal Identity

Robert M. Francescotti, Chair
Department of Philosophy

Thomas Weston
Department of Philosophy

David Armor
Department of Psychology

March 27, 2013
Approval Date

DEDICATION

This thesis is dedicated to Dad for teaching me how to think, and my Mom for teaching me how to dream.

ABSTRACT OF THE THESIS

Emergence and the Problem of Personal Identity
by
David Thomas Jerome Neuburger
Master of Arts in Philosophy
San Diego State University, 2013

Philosophical theories of personal identity often tacitly assume that either the properties which make us Persons are easily divorced from our bodies, or, that our being Persons is one-and-the-same with our being human beings. While there is broad support both scientifically and philosophically for the contention that our being Persons is at least in part contingent on the proper development and functioning of our brains, the lack of consensus as to the manner in which the brain and mind relate to engender a Person has made the question of personal identity particularly intractable. One theory, called Emergence, posits the mind is engendered by, but not reducible to, the actions of the brain. In this thesis, the consequences for our conception of personal identity are explored given an Emergentist conception of the mind/body relationship. Mainline psychological and bodily continuity theories are rejected in favor of the Systemic Approach, which contends there must exist both psychological and bodily continuity for a Person to remain the same Person over time.

TABLE OF CONTENTS

PAGE

ACKNOWLEDGEMENTS

I would like to acknowledge the faculty and staff of the department of Philosophy at San Diego State University for their patience, insight, and significant contribution in my development as a lover of wisdom. I would also like to acknowledge Dr. Eric Schwitzgebel of the University of California at Riverside for his mentorship. Lastly, I would be remiss if I didn't acknowledge the love and support of my fiance' Mara, who put up with me while I wrote this manuscript.

CHAPTER 1

INTRODUCTION

To answer even the most basic questions about personal identity is a surprisingly difficult task. In part, this is due to the fact that Persons,[1] like everything else in the universe, change over the course of time. Given that the identity relation technically describes two things which are the same, just what is it that stays the same between the human Person who sits down at their kitchen table for breakfast, and gets up twenty minutes later to leave for work, such that we can assert they are one-and-the-same? Even a query as seemingly stupid as *am I the same Person that started this sentence?* does not have an obvious, definite answer. When we try to find one, we discover that while we all seem to have roughly the same idea of what personal identity is in the everyday sense, we diverge significantly with regard to the particulars.

At the heart of the concept of identity are two ideas. First, that entities[2] possess a property (or set of properties) that make them different from everything else in the world. It is this uniqueness which endows an entity with *identity*. Second, in deference to the fact that everything in our world undergoes physical change with the passage of time, it is assumed that there is an acceptable level of transmutation an entity can undergo without affecting its identity. In other words, if everything in the world undergoes change, then the identity of things in the world isn't linked to one particular physical state of matter, but to an acceptable range of physical states. For example, despite the significant physical differences between the ten-year-old David Neuburger and the adult version, I am considered to be the same Person as that child.

---

[1] David Mackie employs the word Person (with a capital P) to mean “person in the strong sense”, or, an entity with psychological attributes of sufficient complexity so as to have the capacity for reason, memory, self-awareness etc. The use of the capital P is to distinguish “Person” in this special sense from “person” in the more common one. I will be making use of Mackie's convention throughout this essay, so when you see Person (with a capital P), I mean it as Mackie does – as “person in the strong sense”.

[2] We'll be treating the words 'object' and 'entity' as synonymous. See Van Inwagen, *Metaphysics*, 22-25 for an account of what does and does not constitute an individual object.

Interestingly, while these two ideas seem perfectly reasonable, their application can lead us to absurd consequences. Imagine someone invents a Star Trek-like transporter.[3] This machine consists of a chamber with a platform on which anything can be placed. An operator can then set the machine to instantly deliver whatever is on the platform to any place they want with the push of a button. It even works for people - you get in the machine in San Diego, the operator pushes a button, and you disappear from the platform and reappear in a similar chamber in Amsterdam. The entity that steps off the Amsterdam platform is in every discernible way the continuation of you. We have no cause to deny the fact that you and she aren't the same Person. Keeping that in mind, think about what happens if the machine malfunctions and you don't disappear from the San Diego platform. There are now two qualitatively identical beings which have equal justification to claim they are the continuation of the entity that stepped onto the platform in San Diego. Which is you?

While scenarios of this sort are difficult to contend with, there exists an obvious rebuttal. Namely, that we don't have transporters and these sorts of sci-fi thought experiment scenarios are extremely uncommon. Given that the scope of the problem evinced by such objections is limited to the finite details of the matter, and that those details don't often interfere with our use of the concept in everyday life, one might be inclined to treat the issue of identity as an amusing but ultimately pedantic one. After all, outside the realm of philosophy or rare psychological cases, our common-sense conception of it serves us well.

While I agree that, for most people most of the time, the issues raised by Star Trek-esque thought experiments can be safely ignored, the issue of personal identity is profoundly important in the philosophical sense and thus not a pedantic one. When we ask the question *what makes me the same over time?*, we are also asking the deeper and most fundamental of questions *what manner of thing am I?*. Otherwise, in failing to address the question of what exactly is being referred to by the word 'I', we have no way of accounting for what must persist over time such that we can say 'I' am one-and-the-same with the the 'I' that just a moment ago began reading this sentence. Moreover, while an entity's being an 'I' is typically regarded as also meaning that entity is a Person, and that all known instances of Persons have

[3] See Parfit, *Reasons and Persons*, 199-201 or Nozick, *Philosophical Explanations*, 41 for similar examples. If you don't like sci-fi thought experiments in philosophy, see Nagel, “Brain Bisection” for an exploration of the issue of personal identity using people who have undergone a corpus colostomy.

been human, being human doesn't necessarily entail being a Person. For example, a corpse is human but typically not considered a Person. We'll explore this issue more in subsequent chapters, but for now what's important is the recognition that:

1. There is a correlation between the way healthy human beings function and the presence of personhood. In humans, personhood appears to be the result of an ongoing process undertaken by human parts. As far as we can tell, when the process stops, the Person is no more.
2. The only known instances of Persons are human beings.
3. Being the same human and being the same Person are not necessarily the same thing.
4. An account of personal identity must encompass an analysis of what a Person is. As such, and in light of (1) – (3), for a theory of personal identity to be tenable it must take into account:
    a. The manner in which physical parts can relate to engender Persons.
    b. The possibility that the Person and the parts which engender her may not be one-and-the-same.
    c. If (4b) is true, under what conditions can the Person persist through time absent some or all of her properly functioning parts.

As this issue has been, among philosophers at least, debated for millenia, there are a great many theories which purport to explain what's required for a Person to be considered the same over time. Ultimately, many of these theories are slightly different manifestations of either the Psychological Continuity Theory (herein PCT), or the Bodily Continuity Theory (herein BCT). The former posits that humans are essentially Persons, and that what is necessary for a Person to be considered the same Person over time is the continuity of her psychological features. The latter argues that humans are nothing more than their physical bodies, and that a human is the same over time if and only if there is continuity of her body. Unfortunately, neither of the aforementioned classes theories of personal identity take into account the possibility of an Emergent relationship between the mind and body. This oversight is to the detriment to the efficacy of each, for if the mind is in fact Emergent from the body, all the theories of personal identity which fall into either category become untenable.

In the following chapters, I will elucidate why the question of what makes a Person the same over time is so difficult to adequately address. I will then give an account of the general theory of Emergence and describe how, if we accept that the relationship between the mind and body is an Emergent one, we are compelled to reject both PCT and BCT. I will then

put forth a new theory of personal identity, explore some of the potential objections to it, and conclude with a summation of my argument.

CHAPTER 2

CHALLENGES ENDEMIC TO IDENTITY

Central to the concept of identity is distinctiveness. If something is indistinct, then there is nothing to differentiate it from the rest of the world as a unique object and thus it has no identity of its own. Distinctiveness comes in two flavors: qualitative and quantitative.[4] The qualitative has to do with the features of the thing – is it red, is it made of metal, is it a gas, etc. The quantitative speaks to the number of that particular thing in existence. For an object to possess identity at any given time, it is sufficient that be qualitatively distinct, but necessary AND sufficient that it be quantitatively distinct. This is because qualitative distinctiveness also makes it quantitatively unique, while a lack of quantitative uniqueness entails the necessary lack of any kind of universally distinguishing feature and thus robs the object of identity.

In this chapter we'll be describing some of the challenges that arise when delving into issues of identity. While it is only quantitative distinctiveness that is necessary to set an entity apart from the rest of the world, we'll be discussing failures of the qualitative variety as well as it serves to elucidate how intractable the problem of identity really is.

2.1 THE BOUNDARY PROBLEM

The boundary problem deals with the qualitative features of an object.[5] Moreover, it's commonly assumed that identity is solely a function of an object's intrinsic properties. That is, that for everything that exists, there is a set of intrinsic characteristics that endow it with distinctiveness – and thus endow it with identity. Parfit, in accord with, and paraphrasing, Bernard Williams, writes “Whether a future Person will be me must depend only on the intrinsic features of the relation between us. It cannot depend on what happens to other

[4] See Noonan, “Identity,” 2 for a good account of this.

[5] Unger, “Conscious Beings,” 312-321.

people."[6] This means that an object's having distinguishing features, and thus having identity, is solely a function of only that object. To put this another way, no matter what happens in the world in which the object exists, it will always be distinct and thus possess identity.[7] It therefore follows that we should be able to resolve questions about an object's identity by virtue of the analysis of the object alone. Unfortunately, in many cases this is not so because in most cases, it's impossible to determine exactly where an object's spatio-temporal boundaries lie.

Barring circumstances akin to a scientist isolating a single elementary particle in a chamber devoid of absolutely everything but that particle, it's impossible to determine exactly which parts belong to an object and which do not. For example, if I were to view the edge of a penny with a microscope capable of showing me both the sub-atomic particles that constitute it and the particles occupying the space around it, by what method would I determine which are a part of the penny and which aren't? Like the drop of rain conjoining the puddle, the two sets co-mingle such that there is no discernible boundary betwixt them.

For the incredulous, I would remind you that outside laboratory conditions, we're not just dealing with a penny. There's also the issue of the dirt, bacteria, and condensation coating the outside to deal with. While at a microscopic level we can distinguish between these things, at the sub-atomic level, there is no qualitative distinction between that which constitutes the penny and the adjacent stuff that doesn't. At that level, it's just a mass of commingling cosmic building blocks - none of which have any qualitative features enabling us to say with objective certainty which belong to the penny and which do not. As such, we have no means of determining what the ontological boundary is between the penny and the rest of the world.

This is in stark contrast to what one would expect. There should be, in all cases, a non-subjective answer to the question *what exactly constitutes the penny and what does not.* As we've just seen, there isn't. To make things even more difficult, the boundary problem

---

[6] Parfit, *Reasons and Persons*, 267. Forbes, "Origin and Identity" also presents an argument rejecting the claim that extrinsic features matter with respect to questions of identity.

[7] There is a glaring flaw in the contention that extrinsic features don't matter (namely that quantitative distinctiveness, a necessary condition for the presence of identity, is not solely a function of the entity in question) which will be discussed in more detail later in the chapter.

applies both physically and temporally. Just as it is often impossible to determine exactly what the physical boundary is denoting what is and is not an object, so too is it difficult to discern a precise temporal boundary between the times when the object doesn't exist and the time in which it does. That is – at what precise moment did the collection of penny-parts transform into a penny? Was it the instant the mint finished pressing the image of Abraham Lincoln into the metal? When exactly did that occur, and why that moment and not the moment just prior? For, it seems that in the case where a mint compresses the metal of one proto-penny by a factor of ten units and the next by a factor of 9.99, we would be perfectly justified in referring to both as pennies. However, if the second is considered a penny, then the first became a penny prior to the mint machine finishing its work – forcing us to acknowledge the fact that the contention *the penny is made the moment the mint is finished pressing the image into the metal* doesn't hold in all cases. It seems as though pennies, and in fact a great many objects in the world, do not have definite boundaries with respect to space and time.

At this point it might be tempting to argue that the issue may not be the lack of definite object boundaries, but our lack of ability to discover them. Unfortunately, as tantalizing as the such an answer might be, it alone is insufficient. Any appeal to the unknowability of an object's boundaries on epistemic grounds must be accompanied by an account of how it's at least metaphysically possible to derive a non-subjective answer to the question. Otherwise, there is no reason beyond faith to induce us to believe such an answer exists.

For example, the process by which we predict the weather is only reasonably accurate and at most only for a week in advance. This is due to the fact that the results of the mathematical models we use to predict the weather are highly susceptible to the accuracy of the data we feed into them. Given that we are incapable of producing perfectly accurate data to feed into our models, we are unable to perfectly predict the weather. Thus the issue is that we are epistemically limited, not that an answer doesn't exist. Returning to our penny example, what happens if we stack one penny atop another? Given the lack of qualitative distinctiveness at the sub-atomic level between the parts belonging to one and the parts belonging to the other, there's simply no means of determining where one ends and the other begins, and no evidence suggesting there is an objective boundary distinguishing them.

Therefore, I contend the boundary problem isn't merely epistemic, it is endemic to the concept of identity.

2.2 The Extrinsic Feature Problem

The extrinsic feature problem deals with both the qualitative and quantitative properties of an object. While the transporter scenario in the introduction is an excellent example of it, let us start with something a bit less sci-fi and come back to it.

Imagine I put three small cubes on a table and align them in a row. Qualitatively, the boxes are identical in every respect except that the first box is painted red while the others are blue. Quantitatively, they are all unique such that we can say the first box has two intrinsic features which endow it with identity whereas the other two have one.

In our box example distinguishing characteristics can be changed to non-distinguishing characteristics without altering the object itself. With respect to color, we need only paint the remaining boxes red. We've made no change to the first object, yet in changing the color of its siblings, we've altered it such that one of the first box's distinguishing features is no longer distinguishing. Moreover, for any feature that qualitatively distinguishes one box from the others, we can affect a change extrinsic to it such that it is no longer distinguishing. This holds true not just for boxes, but for any object, anywhere. Therefore, while non-relative[8] qualitative features are intrinsic to an object, the fact that they are distinguishing is not. As such, whether or not such a feature is relevant to the establishment of its identity is not solely a function of the object itself, but also to the context in which the object exists.

We've seen how extrinsic features to an object can, in part, determine whether or not an object's non-relative qualitative properties are distinguishing - but what about quantitative distinctiveness? Can extrinsic features play a role in defining whether or not an object is quantitatively unique?

Recall that in our transporter example, in possible world [x] a Person steps into the transport chamber in San Diego. The machine is activated and instantly she disappears from the San Diego pad, is converted into energy, sent to a similar device in Amsterdam, and is

[8] An example of a relative qualitative feature would be John being taller than Mary. It's a feature of John's that he's taller than Mary, but it only exists by relative to Mary and thus, isn't completely intrinsic to John.

reconstituted such that in every discernible way she's the continuation of the Person who a moment prior was in San Diego. We are therefore inclined to regard the Person who appears in Amsterdam as the same entity as the Person who disappeared in San Diego. Possible world [y] is exactly the same as [x] with the exception of a malfunction in the transportation device. Instead of disappearing from San Diego, the Person remains AND a qualitatively identical Person (one which we were perfectly content to call the continuation of the pre-transport Person in world [x]) ends up in Amsterdam.

What we have in [x] and [y] is a set of worlds that contain the same entity where in one world she is quantitatively unique and in which the identity relation is confused because the quantitative uniqueness is lost. That is – in world [x], after the transportation, there is only one Person in the whole universe that has her qualitative features. After the transport in world [y], there are two Persons who are continuous with the pre-transport Person. There is no difference between the persons in Amsterdam in worlds [x] and [y] respectively. Yet, despite the fact that the change occurred extrinsically to the Amsterdam entity in world [y], in [x] there is one of her in the world and in [y] there is two. If we are committed to the restriction that only intrinsic features matter, then we are left with two undesirable options.

The first is to deny that the transported entity in world [x] is the same as the entity that disappeared in San Diego. Unfortunately, the only plausible way to accomplish this is to cite the manner of transport (the conversion of a Person into energy and subsequent reassembly) as a break in the chain of psycho-physical continuity.[9] As we'll see in a moment, this runs afoul of the acceptable change problem, and as such, is not a good option. The second is to find cause to deny the identity relation between one of the two resulting entities in world [y] and their pre-split former self. This is equally problematic as you'd either have to cite the break in psycho-physical continuity for the transported entity and contend with the acceptable change problem, or, argue that one of the two resulting entities is somehow less

[9] Williams, "The Self and the Future" argues this point differently - that the possibility of an object's future identity being contingent on features extrinsic to it entails that object losing its identity the moment that possibility comes into play. Forbes, "Origin and Identity" also takes issue with the contention that identity can depend on extrinsic features, but focuses primarily on issues of an object's origins. Regardless, given the fact that it's at least logically possible that a freak cosmic event could occur at any moment such that an exact duplicate of some presently quantitatively unique entity is instantiated, it seems as though Williams is committed to the view (as a consequence of his argument) that nothing has identity at any time.

the continuation of the pre-transport one. So in the case of world [y], our intuition would likely inform us that the San Diego entity is more the continuation of the pre-transport Person than the 'beamed' one (the one which materialized in Amsterdam) because the former has an unbroken chain of psycho-physical continuity while the latter technically doesn't. Robert Nozick, is an advocate of the latter option which he calls 'the closest continuer theory,'[10] which he defines thusly:

> The closest continuer view presents a necessary condition for identity; something at t(2) is not the same entity as x at t(1) if it is not x's closest continuer. And "closest" means closer than all others; if two things at t(2) tie in closeness to x at t(1), then neither is the same entity as x. However, something may be the closest continuer of x without being close enough to it to be x. How close something must be to x to be x, it appears, depends on the kind of entity x is, as do the dimensions along which closeness is measured.[11]

However, being that we're philosophers, we need something more than our intuitions to go on if we are to choose. That is, we need to know by what criteria we are to determine which is closer and which isn't? Nozick sums up the problem thusly:

> Which particular properties, features, and dimensions constitute the measure of closeness, and with what relative weights? The closest continuer theory is merely a schema; what then are its particular contents? What precisely is the metric, why that one, and why is it precisely that which we care about? Does psychological continuity come lexically first; is there no tradeoff between the slightest loss in psychological continuity and the greatest gain in bodily continuity; is bodily continuity (to a certain degree) a necessary component of identity through time; how are psychological similarity and bodily similarity to be weighed (for noncontinuers when some other continuer is present); what are the relevant subcomponents of psychological continuity or similarity (for example, plans, ambitions, hobbies, preferences in flavors of ice cream, moral principles) and what relative weights are these to be given in measuring closeness? (Ibid., 69)

Unfortunately, he never answers his own question. In the following paragraph that he states that he will:

> ...make no attempt here to fill in the details; and not merely because (though it is true that) I have nothing especially illuminating to say about the details. I do not believe there are fixed details to be filled in; I do not believe there is some one metric space in which to measure closeness for each of our identities. (Ibid.)

---

[10] Wiggins, *Sameness and Substance*, and Heller, "Approach to Diachronic Identity," both opponents of the view, refer to it as the 'best candidate approach'

[11] Nozick, *Philosophical Explanations,* 34.

Clearly, this is problematic if we are to take the closest continuer theory seriously. Even if Nozick is correct in his claim that there isn't one schema capable of serving as a means of measuring closeness, there's no reason to assume there aren't several – each individually appropriate to different circumstances. Further, and perhaps more importantly, theories are only useful if they answer more questions than they create. Absent any sort of criteria by which we are to determine which among a set of possible candidates is the so-called closest continuer, this theory is of no use to us because it gets us no closer to contending with problematic cases involving identity. Finally, this issue isn't limited to the closest continuer theory, but to every theory of identity. Given that everything in the world changes over time, we have to know which changes matter with respect to identity and which do not. This is germane not only to the determination of the closest continuer, but also in our understanding of what must persist over time in order for the identity of an object to be maintained.

2.3 The Acceptable Change Problem

As stated in the introduction, one of the underlying concepts of identity is that because everything in the physical universe undergoes change over time, there is an acceptable degree of transmutation an object can undergo such that if we examine the object at two different times we may reasonably conclude that in both instances we're looking at the same thing. For example, if I were to spontaneously turn into a ham sandwich, it would be pretty obvious that I have undergone a sufficiently significant change such that I am no longer the same entity as I was before. However, if the changes are less radical, and happen over a longer period of time, it becomes impossible to determine exactly when the ontological change occurs. Therefore, and much like what happens when we try to use Nozick's closest continuer theory, statements about the identity of things in the world end up being the result of a fuzzy judgment call regarding what an acceptable amount of change is between entity [x] at time [t1] and entity [y] at time [t2] such that we can say [x] and [y] are the same. It's fuzzy because no matter what we've done, we have yet to determine exactly what amount, and what kind, of change makes an object different or the same.

2.4 Sorites and Ontological Relativity

The issue of change and identity has vexed thinkers for thousands of years. In fact, it was the ancient Greeks who gave us the sorites paradox. Much like identity, the paradox the result of a simple question: *how many grains of sand make a heap?* While everyone would agree that a single grain of sand does not constitute a heap, and that a pile consisting of a substantial quantity of sand does, none of us can answer non-arbitrarily exactly how much sand it takes to make a heap. Is it one hundred grains of sand? Perhaps a thousand? Why that number, and not one less or one more? The qualitative difference between 1,000 grains of sand and 1,001 grains of sand is negligible – so why is one a heap and the other not?

Interestingly, while most reasonable persons would agree that there are some configurations of sand that are heaps and others that aren't, in light of the fact that there is no certainty as to what exactly constitutes a heap and what doesn't, the range of what can and cannot be called a heap is exceptionally wide. Almost every collection of sand can justifiably be referred to as a heap – making the meaningfulness of the word 'heap' almost nil. While this conclusion probably won't bother too many people as it doesn't much matter whether or not heaps exist, but what of the persistence of persons over time? Absent a fact-of-the-matter regarding things like an entity's boundaries or what is and is not acceptable change over time, are we forced to conclude that, like heaps, the distinction between that which is a Person and that which isn't is so vague that the word 'Person' is almost meaningless?

2.5 Problems Unique to Personal Identity

Thus far we have only explored problems endemic to identity in general. It should be noted that, in addition to the challenges we've already enumerated, there are additional issues that are specific to the identity of Persons.

The debate about personal identity centers largely around the question *what matters with respect to the identity of a Person - the mind which makes it one, or the body which engenders it?* With objects like rocks, or toaster ovens, or clouds, there is only the persistence of a particular configuration of matter to reckon with. With Persons, it's entirely possible that the continuation of the body matters very little. Further, and perhaps more importantly, we need to figure out exactly what a 'Person' is.

Generally speaking, an entity is referred to as a Person if he/she/it has a rich mental life in which they are capable of regarding themselves conceptually in both the third and first Person. However, as with the more general problems of identity, there has yet to be discovered an account for personhood that, as a fact-of-the-matter, distinguishes between Persons and non-persons in all cases. For example, while the reader of this manuscript is likely a Person, and the computer on which I've written it is (hopefully) not, what about a dolphin? The definition above could certainly be construed to include dolphins and, in fact, a wide variety of higher-order mammalian life. While I have no problem with idea of dolphin-personhood, it would be nice to know for certain whether or not Douglas Adams was right and dolphins are, in fact, trying to tell us something.[12]

Interestingly, an often ignored consequence of this notion of personhood is that, while all known instances of Persons are human, not all humans are Persons. A fetus does not meet the criteria because it lacks the 'rich mental life' our definition requires. Neither does a human suffering from late-stage Alzheimer's disease or someone in a persistent vegetative state. This despite the fact that they are all living human beings. Moreover, if we accept this definition of personhood, then we must also accept the fact that what we as human Persons identity ourselves as (i.e. as Persons) is not a function our being human, but an attribute of manner in which our human parts relate. Therefore, for an account of personal identity to be adequate guide as to what must persist for a human Person to remain the same Person over time, it must in some fashion account for the manner in which our parts relate such that we can explain why a suffer of late-stage Alzheimer's is not a Person while a normally functioning healthy adult human is.[13]

To date, the lack of an adequate understanding of the manner in which the mind and body relate, or even what a Person is, has made the problem of personal identity particularly intractable. This because our notion of human personhood is entangled with the notion that intangible mental features are somehow engendered by the human brain. Fortunately, a promising and relatively new theory called Emergence may be our salvation. If the

---

[12] Adams, *Thanks for all the Fish*.

[13] It should be noted that the philosophical definition of 'Person' may diverge from a legal one or one based on religious doctrine. It's therefore possible that an entity is a Person in the legal and/or religious sense but not philosophically.

Emergentists are correct and the mind is, in fact, an emergent property of the body, there may yet be a way to answer questions about personal identity in a fact-of-the-matter fashion without running afoul of the challenges enumerated in this chapter. Let us explore Emergence a bit to see how this might be accomplished.

CHAPTER 3

EMERGENCE AND CONWAY'S GAME OF LIFE

Emergence, put simply, is a philosophical concept which describes the manner in which something can be more than the sum of its parts. It purports to explain how novel behavior can arise from, yet be irreducible to, a system's parts.[14] While the scope of the concept covers a wide variety of natural world phenomena, what makes it particularly enticing is the possibility that Emergence is an efficacious model for the manner in which the mind and body relate.[15]

At present, proponents of the idea have yet to come to a consensus as to how exactly Emergence is supposed to work.[16] While there is agreement as to the general idea, there is significant discord as to the specifics. What's presented here is a mainstream conception of it as evinced by Conway's Game of Life.

3.1 ORDO AB CHAO (ORDER FROM CHAOS)

The notion of Emergence is rooted in John Stuart Mill's *System of Logic* (1843) and flourished for about eighty years.[17] It was an attempt to explain how natural-world phenomena like chemical bonding occurred despite the fact that the laws of physics, as they understood them, offered no explanation as to how this occurred. McLaughlin writes:

> And there were, in Mill's Bain's, and Lewes's lifetimes, no even remotely plausible micro-explanations of chemical bonding. The laws concerning which

[14] Where 'system' is defined as a collection of parts relating to one another according to a set of rules.

[15] See Sperry, "Defense of Mentalism"; Hofstadter, *Gödel, Escher, Bach,"* 337-390; or Chalmers, *The Conscious Mind,"* 129-130 for an account of how this might work.

[16] See O'Connor and Yu Wong, "Emergent Properties" or Cunningham, "Reemergence of 'Emergence'" for a survey of the many flavors of Emergence.

[17] While the modern conception of Emergence is rooted in the works of Mill, Broad, Alexander, and others, it is considered to be somewhat distinct from its predecessor. As such, the first iteration of Emergence is often referred to as *British Emergentism*, whereas the more modern form, which came into being in the late 20th century, is typically referred to as just *Emergence*. See McLaughlin, "British Emergentism"; O'Connor, "Emergent Properties"; or O'Connor and Yu Wong, "Emergent Properties" for a history of Emergence as a philosophical theory.

> chemical elements have the power to bond with which others seemed to many like 'brute facts' that admitted of no explanation. To borrow a phrase from Wordsworth which was a favorite of the Emergentists, it was thought that the laws of chemical bonding had to be accepted with 'natural piety.'[18]

When Quantum Mechanics was discovered, so too was an explanation of how things like chemical bonding worked. In Quantum Mechanics, we had a bridge between the laws of physics and the phenomena described by the special sciences. In light of this better model of natural phenomena, British Emergentism was abandoned. McLaughlin explains:

> Quantum Mechanics and the various advances it made possible are arguably what led to British Emergentism's fall. […] Prior to that revolution, Emergentism's main doctrines appeared to many to be exciting empirical hypotheses worthy of serious consideration. Emergentism was mainly inspired by the dramatic advances in chemistry and biology in the nineteenth century that made psychologically salient the conceptual chasms between physics and chemistry and between chemistry and biology, and by the failures of various attempts to build conceptual bridges to cross those chasms. (Ibid., 23, 24-25)

The salient point is that the British Emergentists used measurable, real world phenomena as a guide when forming their theories. Given the knowledge of the time, there was good reason to believe the laws of physics were incapable of elucidating how things like chemical bonding worked. Emergence was a theory which attempted to fill the explanatory void, and was abandoned when what British Emergentists were attempting to model was better understood.

The current interest in Emergence is rooted in the hope that it can explain the manner in which the brain engenders a mind. It's important to note that, beyond the significant advances made in the field of neuroscience between Mill's time and the present, one of the major reasons Emergence is being pursued now and not earlier is the advent of computers. Stephen Wolfram explains: "Undoubtedly, therefore, one of the main reasons that the discoveries I describe in this chapter were not made before the 1980s is just that computer technology did not yet exist powerful enough to do the kinds of exploratory experiments that were needed."[19] The discovery to which Wolfram refers is that 'complexity' can, and often does, arise from systems comprised of very simple parts and rules. Admittedly, he doesn't

---

[18] McLaughlin, "British Emergentism," 23.

[19] Wolfram, *A New Kind of Science,* 46.

refer to this behavior as Emergent, but this may be attributable to a mere difference in professional lexicon. The complexity to which he refers is commonly referred to by philosophers as novelty. As you may recall from a few pages ago, an Emergent system is one that exhibits novel behavior that arises from, yet is irreducible to, the system's basal parts.

It should be noted that there are other features which are commonly associated with Emergence. Francescotti writes "The features widely considered crucial to emergence include *novelty, unpredictability, supervenience, relationality,* and *downward causal influence."*[20] Here a distinction must be made between the first two features and the rest. Whereas novelty, unpredictability, and irreducibility (another feature commonly associated with Emergence not explicitly mentioned by Francescotti) can be categorized as result-properties, supervenience, relationality, and downward causal influence are mechanism-properties. What I mean by this is that mainline Emergent theory posits that the manner in which systems manifest the properties of novelty, unpredictability, and irreducibility is via the mechanisms of supervenience, relationality, and downward causal influence.

While mainstream Emergence considers unpredictability an important result-property, I do not. As Francescotti, O'Connor and Yu Wong, and others have noted, unpredictability is an epistemic issue and not an metaphysical one.[21] O'Connor and Yu Wong write:

> One moral from our previous remarks is that if a notion of emergence is to improve upon the unhappy alternatives rejected above, it had better be robustly ontological. This is contrary to the tendency among recent philosophical commentators to conflate or blur ontological and epistemological issues when applying emergentist ideas to nonlinear phenomena, artificial life, and human mentality. Discussions of senses of 'unpredictability' dominate these other expositions. While seeing how an emergent property would be unpredictable from a certain, limited empirical standpoint is a useful way of getting a fix on the concept, this is but a consequence of its core metaphysical features.[22]

---

[20] Francescotti, "Emergence," 47.

[21] Both Francescotti, "Emergence" and Chalmers, *The Conscious Mind,* 129-130 note that in Broad's conception of Emergence, unpredictability is not merely epistemic, but an essential feature of emergent properties. Given that the brand of Emergence Broad and the other British Emergentist's put forth was shown to be a fallacious model with the advent of Quantum Mechanics, contentions put forth by the British Emergentist school should be treated as suspect. Therefore, absent specific, real-world examples of how an emergent property can be unpredictable, it makes no sense to treat unpredictability as a necessary feature of Emergence.

[22] O' Connor and Yu Wong, "The Metaphysics of Emergence," 66.

Therefore, in light of the fact that our goal is to set criteria for what can and can not be called a system with emergent properties, we are right to limit the domain of acceptable criteria to those features that are metaphysical. Epistemic concerns involve a wider set of issues that are not germane to the question *what sort of thing is an emergent property?* So, if we're going to understand what an emergent property is, we need to determine what is meant by 'novel behavior', and by what standard we are to consider one thing reducible to another.

With regard to novelty, O'Connor and Yu Wong state: "We might say that it [novelty] is 'nonstructural,' in that the occurrence of the property is not in any sense constituted by the occurrence of more fundamental properties and relations of the object's parts."[23] While this definition is non-technical, it is sufficient for our purposes as it adequately conveys what's generally thought of as a novel property.[24] In restricting the domain of novel features to that which is not a property of any of a system's basal constituents or a resultant property (to use Mill's phrasing) of the relations between those constituents, we have effectively defined novelty as a property of an emergent whole which is not reducible to its parts.

As to reduction, John Searle points out that "Most discussions of reductionism are extremely confusing. Reductionism as an ideal seems to have been a feature of positivist philosophy of science, a philosophy now in many respects discredited. However, discussions of reductionism still survive, and the basic intuition that underlies the concept of reductionism seems to be the idea that certain things might be shown to be *nothing but* certain other sorts of things."[25]. In this, as in many things, he is correct. Reduction is confusing. Beyond the five different types of reduction he enumerates which are all forms of the 'nothing but' relation, there is a plethura of published literature on the subject.[26] For our purposes, it will suffice to employ something akin to what Searle expressed as 'the basic

---

[23] O'Connor and Yu Wong, "Emergent Properties," 19.

[24] See Francescotti, "Emergence." 48-50 for a survey of technical and general conceptions of novelty.

[25] Searle, *The Rediscovery of the Mind,* 112.

[26] One interesting example is Jaegwon Kim's (1999) in which he proposes a method of reduction whereby the higher-order properties are put in context of their function, then correlated with whatever properties are instantiated by the set of basal elements. In this manner, higher-order properties can be explained in terms of the underlying basal processes. He calls it 'the causal inheritance principle' which states that if system [s] has novel function [nf] by virtue of [s] having property [R], then the causal powers [nf] have are identical with the causal powers [R] has and thus, [nf] is reducible to [R].

intuition that underlies reductionism'. Let us say that if [x] is reducible to [y], then all of [x]'s features must be accounted for by virtue of analysis of [y] and [y]'s relations *alone*.

To elucidate this interpretation of reduction, and to demonstrate why it is sensible for us to define reduction in this manner, let us evaluate two related statements:

1. [x] engenders [y]
2. [x] is nothing more than [y]

In both cases, [y] would not exist if not for [x]. This does not mean, however, that both statements necessarily entail reduction. We are far more inclined to cite instances of statement (2) as reductive than we are statement (1). After all, in light of the fact that just about everything in the physical universe was made manifest by virtue of some other process, allowing instances of statement (1) to be considered reductive would have the consequence of reducing almost everything to their progenitors. Further, those progenitors themselves have progenitors, who also have progenitors, and so on and so forth. In the final analysis, and depending on your particular worldview, treating engenderment as sufficient condition for reduction has the consequence of making everything everywhere reducible to the Big Bang, God, or perhaps the Flying Spaghetti Monster.

While there's nothing logically wrong with treating statement (1) as a sufficient criteria for reduction, it doesn't serve the intended purpose of the concept.[27] What we want from reductionism is not knowledge of a thing's origins, but the ability to better understand it by accounting for its features and behaviors in a different, somehow advantageous, lexicon. This is particularly useful in the sciences because it enables us to understand, and often predict the behaviors of, complex phenomena in the physical universe. Lightning, for example, is no longer a mystery to us because we now understand lightning to be *nothing but* electrostatic discharges. Moreover, while the cause of these discharges is electrical imbalances in the atmosphere, we wouldn't say that lightning is reducible to the electrical imbalance which engendered it. We rightly recognize the ontological distinction between cause and effect, and as such, typically limit the scope of the meaning of reduction to some form of the 'nothing but' relation.

[27] i.e. it's not logically fallacious to claim reduction encompasses that which engenders an object, just unproductive in the same way using a screwdriver to hammer a nail into the wall is unproductive. It's possible, but there are better tools available to perform the work at hand.

When there is an obvious difference between that which engenders and that which is engendered, settling issues of reduction is typically a straightforward affair. For example, it is clear to us that the statement *lightning is nothing but a discharge of atmospheric electrostatic energy* is true whereas the statement *lightning is nothing but an imbalance of electrical energy in the atmosphere* is not. Clearly, the clauses 'lightning' and 'discharge of atmospheric electrostatic energy' are synonymous in a way that 'lightning' and 'imbalance of electrical energy in the atmosphere' are not. Where things get muddled is when the begat and the begotten coexist synchronically in the same space. For example, is the word "APPLE" nothing more than the letters which comprise it? In light of the fact that both the word "APPLE" and the letters which engender it coexist, does it make sense to treat them as distinct?

As this issue must be reckoned with if we are to create a tenable, standard model of Emergence, let us treat the Apple Question as paradigmatic and, in answering it, elucidate why it is sensible to define reduction as we have.[28]

3.2 The Apple Question

Recall that we have defined reduction thusly: *if* [x] is reducible to [y], *then* all of [x]'s features must be accounted for by virtue of analysis of [y] and [y]'s relations alone. To make this easier to work with, let us restate the claim in terms of the apple question: *if* [the word 'APPLE'] is reducible to [the letters which comprise the word 'APPLE'], *then* all of [the word 'APPLE']'s features must be accounted for by virtue of analysis of [the letters which comprise the word 'APPLE'] and [the letters which comprise the word 'APPLE']'s relations alone.

The first thing we must address is what exactly is meant by the phrase "by virtue of analysis of the letters which comprise the word 'APPLE' and the letters which comprise the

[28] While I make this point throughout the paper, I feel it requires additional emphasis. Reduction is a concept for which there is no fact-of-the-matter definition. There are many logically valid ways to define and employ reductionism. What I present here is one such method that is particularly advantageous with respect to understanding how systems with Emergent properties work. It is not argued that what follows is the only way to define reduction, but that defining reduction in the manner I describe is sensible as it gives us what we want from the concept – a standard by which we can determine whether or not certain things are nothing but certain other sorts of things.

word 'APPLE's relations alone". We can say that the letters which comprise the word 'APPLE' consist of five letter-tokens representing four letter-types of the English alphabet. Each type has unique grammatical, syntactic, and phonic properties. One of the types, the letter 'A', is an article. It's important to note that the letters {A, E, L, P, P} are not, sans their relational properties, one-and-the-same with the word 'APPLE'. To account for the words which the letters engender, we look to both the letters AND the manner in which they relate. Moreover, the phrase "the letters which comprise the word 'APPLE's relations" refers to the entire set of relational properties possessed by all the letters, not just some of them.

In framing reduction in terms of the 'nothing but' relation, we are in effect claiming that when one thing is reducible to another, they are ontologically one-and-the-same. One need only reflect on the usual meaning of the phrase 'nothing but' to see why this is the case. In Apple Question terms, if there are features of the word 'APPLE' that are unaccounted for in the set of properties possessed by the letters which comprise the word 'APPLE' and the letters which comprise the word 'APPLE's relations, then we cannot say the word 'APPLE' is reducible to the letters which comprise the word 'APPLE' and the letters which comprise the word 'APPLE's relations because there is some difference between the two that prevents us from claiming the former is nothing but the latter. Further, if there are relational properties possessed by the letters which comprise the word 'APPLE' that are unrelated to [the letters which comprise the word 'APPLE's role in forming the word 'APPLE', then we cannot say the word 'APPLE' is reducible to the letters which comprise the word 'APPLE' and the letters which comprise the word 'APPLE's relations because, just as in the previous case, there exists some difference between the two which prevents us form claiming one is 'nothing but' the other. In order to make the 'nothing but' relation work, both the thing reduced and that to which the thing is reduced must necessarily be one-and-the-same thing.

Earlier I inquired, *In light of the fact that both the word "APPLE" and the letters which engender it coexist, does it make sense to treat them as distinct?* To this I would say 'no', because within the context of the word 'APPLE', the set of relational properties possessed by the letters themselves give us cause to treat them them not as individual tokens, but as constituents of a complex word-token. That is, while we recognize that the letter 'A' is an individual token, when it manifests itself in the word 'APPLE', it makes sense to treat it not as a discrete element in-and of-itself, but as part of a discrete whole. This is because, in

this context, all of 'A's relational properties pertain to its role in forming the word 'APPLE'. However, what would happen if the letter 'A' simultaneously had more than one set of relational properties?

Imagine you were playing a word game where, given a random selection of letters and a limited amount of time, you had to come up with more words than your opponent in order to win. If you were given the letters {A, E, L, P, P}, you could use the letter 'A' as both a word in- and-of-itself, and as part of the word 'APPLE'. Further, it can be used to form other words, like 'PALE', giving the letter 'A' at least three sets of relational properties. In this context, the word 'APPLE' is *not* reducible to the letters which comprise the word 'APPLE' and the letters which comprise the word 'APPLE's relations because the extra relational properties possessed by the letter 'A' make it fallacious to claim the former is 'nothing but' the latter. In cases where the parts that relate to form a complex whole simultaneously exist as either wholes themselves, or have additional relational properties which cast that part in a different ontological context, the whole is not necessarily reducible to its parts. This despite the fact that a complex token and the simple tokens which engender it may co-exist in the same space at the same time.

As I stated earlier, there's nothing logically unsound about framing reduction in terms of origin – or in the plethora of other ways Searle enumerated. There's no fact-of-the-matter to guide us as to what's correct and what isn't. It is my contention that reduction should be treated as a tool whose purpose is to cast difficult intellectual problems in a more favorable light. As such, we should employ the standard of reduction that is the most efficacious towards that end. Therefore, it makes sense to frame reduction in the manner I elucidated because (1) it preserves the 'nothing but' relation that we intuitively associate with reduction, (2) addresses, in a common-sense fashion, seemingly difficult cases where tokens synchronically engender complex tokens, and (3) gives us a model of reduction that takes seemingly problematic statements like *something be more than the sum of its parts* and makes them comprehensible in a philosophically rigorous way.

In this section I have addressed (broadly and briefly) what features a system must have if we are to regard it as Emergent. By way of demonstration, let us turn to Conway's Game of Life to see an example of how this Emergence stuff works.

3.3 CONWAY'S GAME OF LIFE

Conway's Game of Life was devised in the mid-20th century by Cambridge mathematician John Conway as an instance of cellular automata – a system in which the interplay of a set of simplistic rules and parts can result in complex behavior. Bedau explains how it works:

> This 'game' is 'played' on a two-dimensional rectangular grid of cells, such as a checker board. Time is discrete. A cell's state at any given time is determined by the states of its eight neighboring cells at the preceding moment, according to the birth-death rule: A dead cell becomes alive iff 3 neighbors were just alive, and a living cell dies iff fewer than 2 or more than 3 neighbors were just alive. (Living cells with fewer than two living neighbors die of 'loneliness', those with more than three living neighbors die of 'overcrowding', and a dead cell becomes populated by a living cell if it has the three living neighbors needed to 'breed' a new living cell.)[29]

CGoL, when implemented on a computer, is what Haugeland would classify as an Automatic Formal System. He defines 'automatic' as: "a physical device (such as a machine), with the following characteristics: (1) some of its parts or states are identified as the tokens (in position) of some formal system; and (2) in its normal operation, it automatically manipulates these tokens according to the rules of that system."[30] This is an important feature because at the core of the concept of Emergence is the idea that physical systems can be constituted in such a way that the automatic behavior of the tokens can give rise to novel behavior. As such, we need something which can act automatically for it to be useful in our study of Emergence. He goes on to define 'formal systems' as being capable of 'token manipulation', 'digital', and 'finitely playable' (Ibid., 48).

By token manipulation,[31] Haugeland contends it "means one or more of the following: (1) relocating them; (2) altering them; (3) adding new ones to the position; and/or (4) taking some away" (Ibid., 49). Basically, all that's stated here is that the parts of a system

---

[29] Bedau, "Weak Emergence," 379.

[30] Haugeland, *Artificial Intelligence*, 76.

[31] A system can be generally defined as a collection of parts interacting with each other according to a set of rules. A token is nothing more than one of the system's parts. A complex token is an assemblage of simple tokens which itself is also a token. For example, the English alphabet is comprised of twenty-six letters. These letters are combined to form words. Think of the individual letters as simple tokens, and the words they form as complex. See Haugeland, *Artificial Intelligence*, 48, 71 for more details.

aren't static. This is obviously important as anything we think of in the natural world as possibly having Emergent properties can be defined as a system with non-static parts.

A digital system is defined as "a set of positive and reliable techniques (methods, devices) for producing and reidentifying tokens, or configurations of tokens, from some prespecified set of types" (Ibid., 53). By 'positive' it is meant that "a technique can succeed absolutely, totally, and without qualification" (Ibid.). This matters because one of our ultimate goals in studying Emergence is to better understand the mind/body relationship. Brains can be regarded as systems basally comprised of neural tokens. A neuron can commit only one of two possible actions at any given moment: to fire (discharge electricity) or not to fire. The conditions under which a neuron will fire is positive in that every time a sufficient amount of voltage is fed into it, an action-potential is generated causing the neuron to discharge. This discharge is considered an "all-or-none" action in that the neuron always fires once the threshold voltage has been breached, and the manner in which the neuron fires is always the same regardless of the degree to which the threshold voltage is superseded.[32] Therefore, if one of our goals is to determine whether or not the mind/body relationship is an Emergent one, it's important that our Abstract Emergent System Touchstone be digital.

Lastly, finite playability speaks to the requirement that for any state of a system, the set of valid actions for any token must be derivable. In his words: "In a formal game [system], the procedure(s) for determining whether a proposed move would be legal must always terminate (yes or no); that's the point of finite playability."[33] This is critical if we are to make use of an Abstract Emergent System as our touchstone. Absent the ability to determine, in every instance, whether or not a future state of the system is valid with respect to the rules, we have no means of simulating it in a way that's reliably accurate. Wolfram explains:

> Experience in the traditional sciences might suggest, however, that experiments are somehow always fundamentally imprecise. For when one deals with systems in nature it is normally impossible to set up or measure them with perfect precision – and indeed it can be a challenge even to make a traditional experiment be at all repeatable.

---

[32] Barcroft, "Architecture of Psychological Function."

[33] Haugeland, *Artificial Intelligence,* 69.

> But for the kinds of computer experiments I do in this book [on cellular automata, a type of automatic formal system], there is no such issue. For in almost all cases they involve programs whose rules and initial conditions can be specified with perfect precision – so that they work exactly the same whenever and wherever they are run.…
>
> In addition, looking at systems with simpler underlying structures gives one a better chance of being able to tell what is really responsible for any phenomenon one sees – for there are fewer features that have been put into the system and that could lead one astray.[34]

If we cannot rely on the validity of our model of the system, then we also cannot rely on the validity of the behavior we observe. Given that emergent properties are posited to be the result of systemic behavior, an abstract emergent system lacking the finite playability characteristic is useless to us as a guide to the nature of Emergence.[35]

Now that we understand in broad terms what sort of thing CGoL is, we may ask ourselves *in what way does CGoL exhibit novel behavior that arises from, yet is irreducible to, the system's basal parts?* As Bedau, Rennard, and others have demonstrated, in some configurations of the CGoL grid, simple tokens (individual cells) sometimes self-organize into complex ones. Bedau writes:

> More complex patterns can also be produced by the simple birth-death rule governing individual cells. One simple and striking example – dubbed the 'glider' […] is a pattern of five living cells that cycles through four phases, in the processes moving one cell diagonally across the Life field every four time steps. Some other notable patterns are 'glider guns' – configuration that periodically emit a new glider – and 'eaters' – configurations that destroy any gliders that collide with them. Clusters of glider guns and eaters can function in concert just like AND, OR, NOT, and other logic gates, and these gates can be connected into complicated switching circuits.[36]

And as Rennard states:

> Gliders belong to a class of spaceships, compact mobile patterns that can be used as signals, or information quanta, in collision-based computing devices. Actually,

---

[34] Wolfram, *A New Kind of Science,* 108-109.

[35] It should be noted that there may be random elements like Quantum Indeterminacy that play a role in determining the behavior of emergent systems. As Wolfram, *A New Kind of Science,* 223-296 notes, automatic formal systems like cellular automata are perfectly capable of representing randomness in the basal processes without losing the ability to engender emergent properties.

[36] Bedau, "Weak Emergence," 381.

> spaceships were considered essential for the design of a universal Turing machine and there should be a pattern that produces spaceships.[37]

What we have in these complex tokens is behavior not accountable by virtue of analysis of the simple tokens alone. Recall that a simple token in this case is a single grid on the cell which has only two properties: its state (whether or not it's 'alive' or 'dead') and its relationship to its eight immediate neighbors (which will determine the state of the cell at the next time step). The ability of the complex tokens to act as logic gates and to transmit information is in no way shared by any of basal parts. Therefore, by virtue of our definition of reduction, the self organization of simple tokens into complex ones is an example of novel behavior that arises from, yet is irreducible to, the simple tokens of CGoL.

Recall that earlier in the chapter, we divided the features commonly associated with Emergent properties into two types. As we've just seen in CGoL, Emergent properties are novel and irreducible to the system's simple tokens. But what of the mechanical features of Emergence, supervenience, relationality, and downward causation?

3.4 SUPERVENIENCE

Kim defines Emergence in terms of supervenience thusly: "If property *M* emerges from properties *N(1)..., N(n),* then *M* supervenes on *N(1)...,N(n)*. That is to say, systems that are alike in respect of basal conditions *N(1)...,N(n)* must be alike in respect of their emergent properties." [38] Francescotti points out that "Equivalently, x and y can differ with respect to emergent properties only by differing with respect to the properties of their parts."[39]

This definition holds in CGoL. Given that the complex tokens are constituted by, yet irreducible to, the simple ones, and that the system is of the automatic-formal variety, it's not possible for a change to occur at the complex level without some change in the basal. Further, for any state of the CGoL grid that manifests complex tokens, it does so necessarily. It's simply impossible that, given two instances of the CGoL grid which are identical with respect to the state of the simple tokens, one would manifest complex tokens and the other would not.

---

[37] Rennard, "Game of Life," 3.

[38] Kim, "Core Ideas and Issues," 550.

[39] Francescotti, "Emergence," 50.

3.5 Relationality and Downward Causation

'Relationality' refers to the manner in which the parts of a system relate to one another, and to the system itself. It's important to note that the parts have certain properties only by virtue of their being members of the whole. For example, Francescotti states: "If being alive is emergent, then it supervenes on the properties of an object's proper parts, [...] these properties include those the parts would lack if they were not parts of a living object."[40] A related concept is that of downward causation. Kim defines this as a feature of emergent properties endowing them with "distinctive causal powers of their own, irreducible to the causal powers of their base properties."[41] As these ideas are related in that they both speak to the manner in which the parts and the whole relate, we'll explore them together.

The nature of causal efficacy, much like reduction, can be difficult to make sense of. The general meaning of the term is the capacity to incite some kind of change. In practice, however, it can be quite difficult to determine which thing(s) are acting as causal agents and which are not. For example, imagine a person driving from San Diego to Los Angeles in an automobile. Without the car, the driver would not have the ability to travel at highway speeds. However, without the driver, the car is inert. Which, then, is in possession of the causal efficacy necessary to allow the driver to travel from San Diego to Los Angeles at 70+ miles an hour - the car, or the driver?

Given that both are necessary to achieve the effect,[42] it makes sense to state that both are causally relevant, but in different ways. In this context, the automobile possesses the potential for causal efficacy in that it can act as a mode of rapid transportation, but not the means to express that potential on its own. The driver possesses the capacity to harness the automobile's potential, and cannot accomplish on her own what she could with the car. We can therefore refer to the automobile as a 'causal enabler', and the driver as a 'causal director'. Further, in deference to the fact that change cannot occur *per se*, the laws and features of the

---

[40] Francescotti, "Emergence," 56.

[41] Kim, "Core Ideas and Issues," 557.

[42] A hawk-eyed philosopher may object to my use of the word 'necessary' here because a person could be driving a different sort of vehicle down the highway – a motorcycle for instance. In this case I'm employing the word 'necessary' sans the modal baggage commonly associated with it. What's being posited is not that an automobile is necessary, but simply that a human being typically cannot, without the use of some transportation device with capabilities similar to an automobile, travel at highway speeds.

physical universe can be thought of as a 'causal medium' – that is - a structured environment in which change can occur.[43] Note that every act of change that occurs within a given medium is bound by the rules which govern it. An agent exercising causal efficacy is not creating a change that is somehow outside the bounds of the medium's rules, but simply bringing about some effect, within the scope of what the medium allows, which might not have occurred otherwise.

Let us think of the automobile as a complex token which exists in the causal medium that is the physical world (the sixteen elementary particles which are the purview of quantum mechanics would comprise the set of simple tokens). When a driver enters the vehicle, who is a complex token herself, the two combine to form a more-complex token. The driver starts the car and heads for Los Angeles. Clearly the driver, the automobile, the physical universe, and the combination of driver-and-vehicle are all causally relevant. However, just how do they relate? How do we determine whether or not, as Kim would have us do, the causal powers of one are reducible to the causal powers of another?

As we discovered in our investigation of the Apple Question, there are cases in which an important ontological distinction should be made between the whole and the set of parts which comprise it. The causal power possessed by the car and driver combined is not held by either the car or driver alone. Neither is capable of intelligently navigating the I-5 at highway speeds without the other. As such, because the causal power possessed by the whole is not one-and-the-same with the power possessed by any of the parts, it cannot be reduced.

To further illustrate my point, and because it will be helpful in our discussion of downward causation in CGoL, let us explore what happens if we were to change our thought experiment such that the car is capable of driving itself. That is, without any input from a human other than the selection of a destination, the car is capable of safely propelling itself from San Diego to Los Angeles. Given this change, can we now say that the car, on its own, possesses the same causal powers as the car-and-driver?

A person inclined to believe it does might point out that the benefit of owning a self-driving car is that it's capable of doing everything a human driver can do. It's able to safely

[43] The phrase 'causal medium' is synonymous with the word 'system'. Either term is acceptable as what's being expressed is the notion that change always occurs within the context of a set of tokens whose interplay are governed by a set of rules – e.g. a system.

navigate roads, and deliver passengers to a specified destination. If it weren't so endowed, it would likely be unsafe and nobody would want one. As such, the self-driving car possesses powers beyond that of causal enablement, and beyond that of a typical automobile.

That being said, there is still a key difference between a self-driving car, and a self-driving car with a passenger, which forces us to conclude that their respective causal powers are not one-and-the-same. On its own, a self-driven car can't select its own destination whereas a self-driving car with a person can. That is, if the self-driving car has no passengers and is programmed to drive to Los Angeles, it can't change its mind half-way through and decide it would rather turn around and go to Mexico instead.

Just as we did before, let us think of the self-driving car as a complex token, a passenger in that car as a complex token, and the combination of the two as a more-complex token. As we've discovered, the causal powers of the more-complex token are not reducible to that of its parts. What interesting is that both the complex token that is the car, and the more-complex token that is their combination, have causal powers in their own right. In the same way that the letters which comprise the word 'APPLE', and the word 'APPLE' can coexist in the same place at the same time and maintain their ontological distinctiveness, so too can the causal powers of systemic tokens.

Let us now turn our attention to CGoL and see if relationality and downward causation[44] are properties of an emergent system. In the manner in which Rennard details, let us assume that we've configured a CGoL grid such that it manifests the complex tokens necessary to function as a binary adder. That is, the system functions as a computational device capable of addition.[45] Let us further assume that the system has been setup to calculate a Fibonacci sequence. In other words, the first two numbers fed into the adder are 0 and 1. Thereafter, the first of each input pair is replaced by the sum of the previous two.

Essentially, what we've done is cause the simple tokens of CGoL, an automatic formal system, to engender a second automatic formal system as an emergent property. An automated binary adder calculating a Fibonacci sequence meets the criteria in that it can

---

[44] Given that downward causation has more to do with the lack of reducibility of the causal powers of a complex token to its parts, it might make more sense to rename it to something like "concurrent causation."

[45] Wolfram, *A New Kind of Science,* 656-663 also details ways in which cellular automata can be used as computational systems.

correctly transition from one state to the next on its own, is capable of token manipulation, is digital, and is finitely playable. Just like our self-driven car example, the causal powers of the complex tokens are different than the powers possessed by simple tokens which constitute them. Moreover, the causal powers of the emergent binary adder are also distinct from the basal CGoL system. Therefore, I contend that in the relationship between our binary adder and the basal process that engenders it is an example of downward causation in that it is an emergent property endowed with "distinctive causal powers of [its] own, irreducible to the causal powers of [its] base properties." [46]

Here, relationality works in much the same way. Like a photograph that's been double exposed, our analysis of CGoL has shown us how emergent systems are comprised of multiple sets of tokens (simple and complex) and multiple sets of rules superimposed atop one another. If by virtue of downward causation, the causal powers of emergent properties are distinct from those possessed by the basal constituents, and the tokens which are manipulated by emergent properties are comprised of collections of simple ones, then it stands to reason that the state of the simple tokens is in part informed by the actions of the emergent properties. In other words, the disposition of an emergent system's basal parts is not solely a function of the properties and relationships of and between the parts, but also a function of the fact that those parts are constituents of an emergent whole.

Lastly, the binary adder is a novel feature of this particular configuration of CGoL. Recall that of the standard Emergentist account of novelty, O'Connor and Yu Wong wrote, "We might say that it [novelty] is 'nonstructural,' in that the occurrence of the property is not in any sense constituted by the occurrence of more fundamental properties and relations of the object's parts."[47] They go on to state that "...newness of property, in this sense, entails new primitive causal powers..." (Ibid.). I've already shown how the binary adder has causal powers that are irreducible to those of the basal parts or relations. Let us now dig into why it's novel.

The capacity of this system to act as a binary adder is not reducible to either the properties of CGoL's tokens or the relations between them. The individual cells have only

[46] Kim, "Core Ideas and Issues," 557.

[47] O'Connor and Yu Wong, "Emergent Properties," 19.

one property – life and death. The relations between them are comprised solely of a cell's state being governed by the previous state of the surrounding eight cells. Conversely, the binary adder is comprised of parts that, while engendered by CGoL's tokens and relations, possess wholly unique properties and relations. These parts include logic gates, gliders, and eaters –all of which exist and relate in a manner radically different than their constituent basal elements. For example, unlike the property of the number of living cells on the grid at a particular iteration, in which there is a direct correlation between the properties/relations of the individual cells and this aspect of the whole system, the binary adder shares no such connection with the elements of CGoL. What we have in our binary adder is a second, wholly distinct, system superimposed atop of CGoL. It's comprised of parts which possess properties significantly distinct from those of CGoL, and operates according to a set of rules that is equally disparate. As such the binary adder is not merely constituted by CGoL, it is a distinct system engendered by it, and thus, is novel.

3.6 A General Theory of Emergence

By virtue of our analysis of CGoL, we can now formulate a data-driven hypothesis of what constitutes Emergence. We've concluded:

1. a system is a collection of parts relating to one another according to a set of rules
2. emergent properties are novel features that arise from, yet are irreducible to, the system's basal parts.
3. systems which engender emergent properties have the following features: supervenience, relationality, downward causal influence.

CHAPTER 4

IDENTITY AND THE EMERGENT MIND

4.1 THE NATURE OF IDENTITY

Let us begin by briefly reviewing what it is that endows an entity with identity. First, it must remain sufficiently unchanged over time such that we can claim the object has retained its identity. That is, our means of determining the identity of an object is in part informed by the degree to which it remains self-identical with previous iterations of it. The fact that I am considered the same Person as the ten-year-old David Neuburger is an example of this. Second, an entity must be unique with respect to the rest of the world. While it is sufficient that this uniqueness be qualitative, it is both sufficient AND necessary the entity be quantitatively unique.[48]

In more technical terms, for [x] at time [t1] to be identical with [y] at time [t2], there must exist an unbroken chain of quantitative uniqueness between [x] and [y]. There also has to be chain of qualitative continuity between [x] and [y] such that whatever changes that occurred between [t1] and [t2] are not so drastic so as to give us cause to claim [x] and [y] are dissimilar entities. While it is true that if [x] at [t1] is an acorn and [y] and [t2] is a large oak tree, there is a significant qualitative difference betwixt them,. However, so long as between [t1] and [t2] there are a sufficient number of intermediate interations where the qualitative differences aren't drastic and at no point did the entity become quantitatively non-unique, we may assert the existence of a chain of continuity between [t1] and [t2]. If both criteria have been met, we have sufficient cause to claim [x] and [y] are the same. In essence, identity is a function of the degree to which an object remains the same over time and the fact that an object is unique with respect to the rest of the universe.

[48] This with the caveat that the determination of uniqueness is with respect to an entity at a particular time because of the aforementioned allowance for objects to undergo qualitative change over time and maintain the identity relation.

4.2 The Consequences of an Emergent Mind for Identity

If we accept that the mind is an emergent property of the body in the manner described in chapter three, then in some important respects, the mind of a human being is similar to both cellular automata and colonies of ants.[49] As in CGoL and ant colonies, the neurons which form the basis of the human mind act unwittingly in concert to form an emergent system greater than the sum of the parts.[50] The emergent mind is therefore novel, and both the mind and its basal neurons are systems of actors which apply influence on the other. Finally, as is the case with any emergent system, the mind and the body from which it manifests are neither wholly distinct nor identical, but two systems that, due to supervenience and downward causation, are *ontologically linked*.

To better understand the nature of this ontological linkage, recall that a system is a collection of parts interacting with one another according to a set of rules.[51] In this respect, the body is a system, as is the mind. Therefore, what is essential to the nature of both the body and the mind is their constituent parts and rules. That being the case, there are some consequences of acceptance of the afore described Emergentist account of the relationship between the mind and body. First, one must also accept that a subset of the rules which govern either system are present only by virtue of the connection each has with that to which it is emergently-entangled. Second, two emergently-entangled systems like a mind and body can not be separated without radical change to both. This is because a consequence of the act of separation is the sudden removal of properties essential to the nature of both, namely, the aforementioned subset of rules governing the interplay of system's (the combined mind/body system's) parts. Third, an account of the identity of emergently-entangled systems (the mind

---

[49] Hofstadter, *Gödel, Escher, Bach,* 337-365.

[50] A neuron may commit only one of two possible actions at any given moment: to fire (discharge electricity) or not to fire. It, like the ant, is unaware of its role in the larger system and yet is a basal constituent of it.

[51] The term 'rules' can be defined as laws that govern the actions of the system's parts. In the physical world, the rules are the laws of nature. In CGoL, the rules are the the laws governing whether or not a cell is on or off at a particular time. 'Rules' doesn't simply refer the laws of nature, but to whatever rules are in play that govern the actions of the tokens within the scope of a particular system.

and the body) over time must take the features of both systems into account when establishing qualitative similarity between them.

Taking these consequences into account, with respect to a normally functioning adult human, if the relationship between her mind and body is an emergent one, what is essential to her identity over time as a Person cannot be either the persistence of her mind or body alone because neither exists independently of the other in the same fashion as it does together. However, this is precisely what adherents to PCT and BCT adherents contend *is* possible. PCT posits that we are essentially our minds. BCT contends our persistence over time is contingent on the persistence of our human bodies. In any case, both give primacy to only a part of our essential nature, and in so doing, fail as viable accounts of personal identity as they neglect the consequences of an emergent mind/body relationship. Therefore, in order to address the shortcomings of PCT and BCT, I propose the adoption of a new theory of personal identity.

4.3 The Systemic Approach

As stated in the previous section, if we accept that the mind is an emergent property of the body, we must also accept that a theory of personal identity over time cannot ignore either psychological or bodily continuity because both the mind and body are ontologically linked. Moreover, another consequence of the mind/body relationship being emergent is that, despite the fact that the body engenders the mind, the mind is not reducible to the body as they are ontologically distinct. I therefore argue that a viable theory of the identity of Person engendered by an organism like a human being must take into account the systemic continuity over time of both the mind and body. Understand that what's being posited is not that human beings are essentially Persons, but that normally functioning human beings of sufficient maturity are emergently entangled psycho-physical systems of which Personhood is a property. This is an important distinction because it's at least logically possible that non-human Persons do or will exist. As such, it's imperative that a distinction be made between a human being, and a human Person. We are not fundamentally Persons who happen to be psycho-physical systems. We are psycho-physical systems which sometimes happen to be Persons. As such, a tenable theory of the identity of human Persons over time must have as

its central criteria the requirement for the persistence of all the structures and processes necessary for a human Person to remain the same Person over time.

Because our claim is contingent on the features of an emergent property, and on the emergent nature of the relationship between the mind and body, let us call it the *Systemic Approach* (herein SA).

The theory, more formally, would look something like this:

IF

> A “SYSTEM” IS DEFINED AS A COLLECTION OF PARTS INTERACTING WITH ONE ANOTHER ACCORDING TO A SET OF RULES.

AND

> “SYSTEMIC CONTINUITY” IS DEFINED AS THE CHAIN OF CONTINUITY OVER TIME OF A SUFFICIENT NUMBER OF A SYSTEM'S PARTS AND RULES; WITH THE CAVEAT THAT WHAT PERSISTS NEED ONLY BE QUALITATIVELY EQUIVALENT.

AND

> BODIES AND MINDS ARE CONSIDERED TO BE ESSENTIALLY SYSTEMS.

AND

> [x] AND [y] ARE TEMPOROSPATIALLY DISTRIBUTED ENTITIES THAT
>
> 1. HAVE BODIES
> 2. HAVE MINDS OF SUFFICIENT COMPLEXITY FOR US TO MAKE THE CLAIM THAT THE ENTITY IS A PERSON (WITH A CAPITAL 'P')
> 3. ARE CONFIGURED SUCH THAT THE MIND IS EMERGENT FROM THE BODY

THEN

> ENTITY [x] AT TIME [t1] IS IDENTICAL WITH ENTITY [y] AT TIME [t2] IFF THERE IS SYSTEMIC CONTINUITY IN BOTH THE PHYSICAL AND PSYCHOLOGICAL FEATURES BETWEEN [x] AND [y].

Let us begin by discussing systemic continuity and the caveat that what persists need not be the exact same thing, only qualitatively equivalent. As with any human Person, and speaking from a purely bodily perspective for a moment, you have undergone significant change since you were a child. Your physical appearance and dimensions are radically different now than they were then. The cells which made up your body as a child have all

died and been replaced with new ones. As Joseph Butler was correct to point out,[52] in the strictest philosophical sense you are not identical with that child as there is no part of you today that is truly the same as any part you had then. Yet, despite these significant changes, we are willing to say you are the same entity as that child. Why? *Systemic Continuity*.

A living body is essentially a system, and for us to consider that system to have persisted as the same thing over time, we require a chain of continuity, not between identical parts and rules, but qualitatively equivalent ones. Otherwise, once a sufficient number of cells had died off and been replaced, we would be forced to concede there no longer exists identity between the entity comprised of new cells and the entity comprised of the old ones. The same holds true for the mind. Given that the mind is also essentially a system, the same conditions for identity hold. When we say something is psychologically continuous, we are not claiming that every part and rule which comprises the mental system is identical, only that there exists systemic continuity.

Now a word on what is meant by 'qualitatively equivalent'. In the same way 2+2 and 3+1 are equivalent in their effect but distinct in their manner of achieving it, I mean 'qualitatively equivalent' to encompass all things which may be ontologically distinct yet in effect, equivalent. For example, imagine a Person who had her arm amputated. The doctors give the patient two options. The more expensive option is to use the patient's DNA to grow a new arm in a vat that will be an exact biological duplicate of the one the patient lost. The cheaper option is to use the patient's DNA to build a mechanical arm that, down to the cellular level, will feel and act exactly like the one the patient lost. Despite the fact that one arm is an exact copy of the patient's wayward limb and the other is only a functional duplicate, both are considered to be qualitatively equivalent. Moreover, in terms of the Systemic Approach, any systemic aspect can be replaced with something(s) which will net the same causal consequence without the loss of systemic continuity.

With that in mind, let us address the case where there exists a systemic discontinuity in either the mind or body between two temporospatially distributed entities. In the case of the appearance or disappearance of continuity of either the body or the mind, the other is still considered congruent, but they aren't considered the same Person. For example, if we

[52] Butler, “Of Personal Identity,” 101.

imagine a Person who suffers a traumatic brain injury such that their body lives but her brain is damaged so as to eliminate her higher order mental functions. While there exists bodily continuity between the pre and post accident entity, there is a break in psychological continuity such that the pre and post accident entity are not the same Person. They are merely the same human.

This is also the case with an argument Eric Olson puts forth – that a consequence of the acceptance of PCT is that it obligates us to make the claim that the fetus which grew in your mother's womb, and with which you appear to share physical continuity, is not you. This is because, at the time of conception, a fertilized human embryo has no mental states. Assuming a normal and healthy development occurs, at some point between conception and adult-hood, the entity matures into a Person. Olson argues that an absurd consequence of PCT, which holds that the identity of a Person is solely a function of the persistence of its mental states, is that no adult human is the same entity as the fetus from which they came. He then argues that PCT should be rejected in favor of BCT. However, there is a difference between the claim that the adult is not the same Person as the fetus, and the claim that the two are not the same human. The Systemic Approach allows us to make that distinction such that we can answer's Olson's question by denying you and the fetus from whence you came are the same Person, but affirming that you are the same physical entity.

Conversely, imagine a Person who is ridiculously rich, very bored, and a bit of a sociopath. The technology exists to swap his mind with that of another unsuspecting human. Despite the presence of psychological continuity between the pre-swap entity and the entity into which the swap was made, we cannot claim that the two are the same Person. Let us use John Locke's Prince and Cobbler thought-experiment to examine the issue more closely:

> For should the soul of the prince, carrying with it the consciousness of the prince's past life, enter and inform the body of a cobbler, as soon as deserted by his own soul, every one sees he would be the same Person with the prince, accountable only for the prince's actions....[53]

Locke would have us believe that the Prince in the body of the cobbler is still the Prince and that therefore, personal identity must be a function of the continuance of psychological features. However, by virtue of the fact that the psychological and physical

[53] Locke, "Of Identity and Diversity," 44.

features of a Person are not wholly distinct from one another, there is insufficient systemic continuity between the Prince and the Prince-in-the-Cobbler's-body to claim the two are the same Person. This is because of the break in bodily continuity, which in turn as a significant impact on the properties of the Person. If we accept that the mind is emergent from the body, then we must also accept that the mind and body are ontologically linked such that, by virtue of supervenience, downward causation, and relationality, any significant change in one necessarily coincides with a significant change in the other.

Recall that the human body is essentially a system, that is, it is a collection of parts interacting with one another according to a set of rules. The mind, emergent from the body, is also a system of parts obeying a set of rules. Due to the afore stated ontological linkage between the mind and the body, the rules that govern the system that is the body are partly dictated by the system that is the mind, and visa versa. If what is essential to a system's identity over time is systemic continuity, and, if in a system with emergent properties, the rules by which the basal properties interact are in part informed by the actions of the emergent and visa versa, then, the sudden appearance, absence, or significant change in either the mind or body also entails a significant change other.

To further illustrate why psychological continuity is insufficient to sustain Personal identity, imagine a case where you were in an accident and your body is near death. Your brain is intact. The doctors, unable to save your life in any other way, scoop out your brains and plunk them into a shiny new robotic body. You still have all the memories and psychological what-not that you had prior to the accident, but because there has been a radical change in your body, and in light of the previously established fact that the mind and body are inextricably linked, the pre-accident entity is not the same Person as the post-accident cyborg. For example, whereas the pre-accident entity was interested in amorous relations with other persons, the post-accident entity has no genitals and therefore has no sex drive. It also doesn't have any digestive organs and therefore has no need to eat (it's an eco-friendly hybrid that runs on a hydrogen-oxygen fuel cell). Despite the fact that the cyborg still has memories of eating and making sweet love, it no longer is able to and no longer has any biological imperative compelling it to. The break in bodily continuity necessarily entails a break in Personal identity.

4.4 Possible Objections to the Systemic Approach

The central claim of PCT is that we are essentially Persons and that only psychological continuity is required for a Person to remain the same over time. A PCT theorist would likely point out that, when we reflect upon what is essential to our distinctiveness as Persons, what seems to us to be fundamental is our mental features. We can imagine cases where we lose limbs, get smurfed (a mad scientist turns our skin blue), have organs replaced, etc., and still see ourselves as the same Person because there exists psychological continuity. Therefore, is it logical to claim that what's necessary for the identity over time of a Person is the continuity of her mental states? Further, to contend the Olson fetus argument, PCT theorist might allow that the fetus which grew in my mother's womb engendered me in the same way a caterpillar engenders a butterfly. That would certainly answers awkward questions like: *if that newborn*[54] *that got circumcised isn't me, then when and how did I get snipped?* and *if I was never a fetus, and the fetus which I never was is contiguous with the egg my father fertilized whilst engaged in sexual congress with my mother, then who are my parents?*

The primary issue with PCT is that is seems to assume that psychological features are distinct and somehow unaffected by the manner in which they are manifested. Perhaps if the relationship between the mind and body was akin to that of a driver and an automobile I could accept the primacy of our psychological features as a determinant of personal identity. If the nature of the interplay between the two was of this sort, then there would exist clear, natural boundaries delineating where the body ended and the mind began. The essential 'us' would be wholly distinct from, yet resident in, our bodies. This, after all, is the root of our intuition when we reflect upon cases of our continuity with a physically dissimilar entity. We imagine that, in the same way a Person might exit one car, get into another, and still be the same Person, we can modify or even be transplanted from our bodies without significant consequence to our psychological makeup. After all, if our bodies are extraneous parts, then any change in those parts should not affect our identity.

---

[54] Assuming, of course, that a newborn and a fetus are similar in that they lack sufficient psychological complexity to be considered Persons

Despite our intuition on the matter, our mind and body are not distinct entities, but intimately linked. While the previous thought experiment showed how psychological continuity may be preserved absent the standard notion of bodily continuity, it did not elucidate the manner in which there may be psychological continuity without the existence of systemic continuity of the basal components from which the mind is emergent. Therefore, the claim that what is essential to the identity of a Person is *only* the continuity of their psychological features is erroneous because in making any claim about psychological continuity, one is also making an implicit claim that there is systemic continuity of the mind's basal properties as well. Unless PCT can demonstrate the manner in which psychological continuity can exist absent systemic continuity of its basal properties, I contend it must be rejected.

Derek Parfit contends that any criterion of personal identity cannot hold because it cannot meet Williams' requirement that "Whether or not a future person will be me must depend only on the intrinsic features of the relation between us. It cannot depend on what happens to other people."[55] Here I must emphatically disagree with both Parfit and Williams. If identity were solely a function of the degree to which two temporally displaced entities are similar then Williams' argument would hold. Unfortunately, that's not all there is to identity. An object must also be distinct in order to possess identity, and that distinctiveness necessarily involves features extrinsic to it. This is because, as I elucidated in chapter two, the criteria by which distinctiveness is determined involves the juxtaposition of the thing in question against the rest of the physical universe. Only by comparing an object's intrinsic features against the entire set of features extrinsic to it can one determine whether or not an object is distinct, and thus, meets one of the two criteria for having identity. As such, I contend Parfit was wrong to claim any criterion of personal identity cannot hold true.

[55] Parfit, *Reasons and Persons,* 267.

CHAPTER 5

CONCLUSION

The problems of boundary, acceptable change, and extrinsic features all elucidate the manner in our current conception of identity is lacking. As we have seen, ignoring extrinsic features is problematic for the tenability of a theory of identity due to the ease with which a detractor can construct a scenario in which the theory doesn't hold true. While including the temporal features of an object may be one way to avoid these kinds of issues, it is a yet unanswered question as to whether or not the nature of the world and the objects in it justify the acceptance of such a view.

With a basic framework of what identity entails in a general sense, we can now turn our attention to issues of Personal Identity. Persons, being entities with sufficiently complex and rich mental lives that regard themselves as subjects, have unique features which make answering questions about their identity particularly challenging. Does the persistence of a Person over time entail the persistence of their bodily features as the BCT theorists suggest, or the continuance of their psychological features, as the PCT theorists contend?

It is a common and longstanding misconception among the philosophical community that, just because it is a simple matter to imagine a case where the mind and body become easily divorced, it is possible to do so. As we have seen, acceptance of an Emergentist conception of the mind/body relationship leads us to conclude this is far from the case. As such, the ramifications for both Psychological and Bodily Continuity Theories of personal identity are profound. If the mind of a human Person is Emergent from the body, then neither BCT nor PCT is tenable. Therefore, I contend that in order for a theory of personal identity to be sound, it must take into account the nature of the emergent relationship between the mind and body. The Systemic Approach allows us to do just that, and as such, is humbly put forth as a replacement for its less efficacious peers.

WORKS CITED

Adams, Douglas. *So Long, and Thanks for all the Fish*. New York: Random House, 2005.

Barcroft, Joseph. "Features in the Architecture of Psychological Function." *Journal of Nervous and Mental Disease* 82, no. 3 (1935): 250.

Bedau, Mark A. "Weak Emergence." In *Philosophical Perspectives: Mind, Causation, and World*, edited by James E. Tomberlin, 375-399. Malden, MA: Blackwell, 1997.

Butler, Joseph. "Of Personal Identity." In *Personal Identity*, edited by John Perry, 99-105. Berkeley: University of California Press, 2008.

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

Cunningham, Bryon. "The Reemergence of 'Emergence'." *Philosophy of Science* 68, no. 3 (2001): S63-S75.

Forbes, Graeme. "Origin and Identity." *Philosophical Studies* 37, no. 4 (1980): 353-62.

Francescotti, Robert. "Emergence." *Erkenntnis* 67, no. 1 (2007): 47-63.

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

Heller, Mark. "The Best Candidate Approach to Diachronic Identity." *Australasian Journal of Philosophy* 65, no. 4 (1987): 434-451.

Hofstadter, Douglas R. *Gödel, Escher, Bach: An Eternal Golden Braid.* New York: Basic Books, 1999.

Hume, David. "Of Personal Identity." In *Personal Identity*, edited by John Perry, 161-176. Berkeley: University of California Press, 2008.

Kim, Jaegwon. "Emergence: Core Ideas and Issues." *Synthese* 151, no. 3 (2006): 547-559.

———. "Making Sense of Emergence." *Philosophical Studies* 95, no. 1-2: (1999): 3-36.

Kripke, Saul A. *Naming and Necessity*. Cambridge, MA: Harvard University Press, 1980.

Locke, John. "Of Identity and Diversity." In *Personal Identity*, edited by John Perry, 33-52. Berkeley: University of California Press, 2008.

Mackie, David. "Personal Identity and Dead People." *Philosophical Studies* 95, no. 3 (1999): 219-242.

McLaughlin, Brian P. "The Rise and Fall of British Emergentism." In *Emergence or Reduction?: Prospects for Nonreductive Physicalism*, edited by Ansgar Beckermann, Hans Flohr, and Jaegwon Kim, 19-59. Berlin: De Gruyter, 1992.

Nagel, Thomas. "Brain Bisection and the Unity of Consciousness." *Synthese* 22 (May 1971): 396-413.

Noonan, Harold. “Identity.” In *Stanford Encyclopedia of Philosophy.* Stanford University, 2011-. Article published December 15, 2004. http://plato.stanford.edu/archives/win2011/entries/ identity/.

Nozick, Robert. *Philosophical Explanations*. Cambridge, MA: Harvard University Press, 1981.

O'Connor, Timothy. “Emergent Properties.” *American Philosophical Quarterly* 31 (1994): 91-104.

O'Connor, Timothy, and Hong Yu Wong. “Emergent Properties.” In *Stanford Encyclopedia of Philosophy.* Stanford University, 2012-. Article published September 24, 2002. http://plato.stanford.edu/archives/spr2012/entries/properties-emergent/.

———. “The Metaphysics of Emergence.” *Noûs* 39, no. 4 (2005): 658-678.

Olson, Eric T. “Was I Ever a Fetus?” *Philosophy and Phenomenological Research* 57, no. 1 (1997): 95-110.

Parfit, Derek. *Reasons and Persons*. Oxford, UK: Oxford University Press, 1984.

Rennard, Jean-Philippe. “Implementation of Logical Functions in the Game of Life.” In *Collision-Based Computing,* edited by Andrew Adamatzky, 491-511. London: Springer, 2002.

Searle, John R. *The Rediscovery of the Mind*. Cambridge, MA: MIT Press. 1992.

Sperry, Roger W. “In Defense of Mentalism and Emergent Interaction.” *Journal of Mind and Behavior* 12, no. 2 (1991): 221-246.

Unger, Peter K. “Conscious Beings in a Gradual World.” *Midwest Studies in Philosophy* 12, no. 1 (1988): 287-333.

Van Inwagen, Peter. *Metaphysics.* Boulder, CO: Westview Press, 1993.

Wiggins, David. *Sameness and Substance*. Cambridge, MA: Harvard University Press, 1980.

Williams, Bernard. “The Self and the Future.” In *Personal Identity,* edited by John Perry, 179-198. Berkeley: University of California Press, 2008.

Wolfram, Stephen. *A New Kind of Science*. Champaign, IL: Wolfram Media, 2002.