THE ETHICAL IMPLICATION OF THE EXTENDED MIND THESIS FOR ADVANCE DIRECTIVES AND PROXY DECISION MAKING

By

Michael P. Leeds

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Philosophy—Doctor of Philosophy

2018

ABSTRACT

THE ETHICAL IMPLICATION OF THE EXTENDED MIND THESIS FOR ADVANCE DIRECTIVES AND PROXY DECISION MAKING

By

Michael Patrick Leeds

What one believes about the mind has implications for what one believes about ethics. In this dissertation I consider the plausibility of the extended mind thesis and what the consequences of this theory are for the implementation of advance directives and proxy decisions in a medical setting. I conclude that the extended mind thesis, if true, does mush to resolve longstanding debates over making medical decisions for incompetent persons.

For my wife, Lauren, and our son Henry George Douglas … and my friend Wayne (I guess).

ACKNOWLEDGEMENTS

As with any project approaching this scale I could not have done this without the help and encouragement of friends, family and colleagues.

First, I would like to acknowledge and thank my dissertation committee members both present and past: To my current (and final) committee, Dr. Jamie Nelson (chair), Dr. Robyn Bluhm, Dr. Frederick Rauscher and Dr. Paul Thompson, and my past committee members Dr. Tamara Frei, Dr. Fred Gifford, Dr. Matthew McKeon and Dr. Jennifer Susse (former chair). I thank you for your patience and indulgence. Your feedback and criticism were indispensable in writing this dissertation. You kept this most scattered-brained of people honest and on track. I am a better philosopher for having known you.

But no one's education starts with a dissertation committee. I have had teachers throughout my life who pushed me to learn despite my protestations. Thank you to Ms. Muir (3rd grade) Mrs. Dunklau (4th and 5th grade), Mr. Smith (7th grade English), Mr. Leskey (7th grade law), Mrs. Sage (sophomore English), Mr. Giesler (high school history and sociology), Ms. Nork (junior English) Mr. Guay (high school theater), Dr. Paul Graves (Oakland University, metaphysics and logic), Dr. Elyssa Koppleman-White (Oakland University, ethics) and Walter Langlois (Oakland University, Japanese language and culture). I am a better student and teacher for having known you.

Also, my mom and dad. I literally would not *be* a person without you. Sorry I couldn't get this done fast enough for you to see this, Pops.

My friends, (in no particular order, so no whining) Wayne Smith, Michael Zois and Tracy Murrell, Dave and Jen Foster, Matt and Mari Fox, William Herman Michael Gordon Raul

Tiberius Binder, Scott and Tracy Powell, Jason and Cheryl Ingram, Zachary Blount, Eric Richmond and Jessica Reiss, Abby Rachel Seigel Kelly, Joe Kelly, Carrie-May and Jason Larson, Mark and Jeni Balawender, and Eric Berling. It's over 30 years, in some cases and you still put up with me. You folks are saints.

Importantly, there are two people I would like to give special thanks to:

Bill Isanhart. You were my first philosophy teacher (so, either you get the credit for this, or this is all your fault). You were the first teacher to take my nonsense seriously. You showed me doubt was acceptable intellectual position and, in some cases, the most honest. I am a better teacher for having known you.

And Jorma Leavy. It wasn't until you were gone that I realized what a monumental influence you were on my life. Two months before we lost you, when my son was barely over a month old, I told him "I only want three things for you; be kind, be happy and be your own person, just like Jorma." I'm sad he will never meet you but be assured you will be a presence in his life. I am a better person for having known you.

… Oh yeah, and no thanks to Einstein the dog.

TABLE OF CONTENTS

LIST OF FIGURES

Introduction

"The mind is not a book, to be opened at will and examined at leisure. Thoughts are not etched on the inside of skulls, to be perused by an invader. The mind is a complex and many-layered thing".

Severus Snape

Concepts such as mind, person or self are central to all but a handful of ethical theories. This is both trivially and importantly true. It is trivially true in the sense that if there were no minds, persons, or selves there would be no things capable of engaging in ethical discourse nor would there be beings capable of agency and action, one of the chief subjects of ethics. This is an importantly true claim in that what theories (or even intuitions) one holds about minds, persons, or selves, may not be consistent with certain ethical theories. For example, if one is a solipsist it is apparent that the Kantian claim that we should treat others as ends and not as means is a morally vacuous statement since the solipsist does not believe in the existence of other persons. If one believes all and only humans beings have minds there are implications for how we can treat nonhuman animals, morally speaking. Insofar as one believes the stronger claims about dissociative identity disorder (previously known as multiple personality disorder) or severe retrograde amnesia one may be inclined to show leniency in a court of law because the actions of the defendant were not truly her own. (Taylor, 1984, Nakic, 2012)[1]

[1] The 'inclination' I am referring to relates to Lockean memory theory in that if one does not remember something which actually happened, that happening is not part of that what makes that person who she is. Such mental disorders would make the accused innocent of the charge as she was not the person who committed the egregious act. That being said, defenses of the form "my client doesn't remember doing it, therefore you have the wrong guy" have been tried in American courts for decades. Legally, however, this has never been found to be, in and of itself, an adequate defense outside of times when the lack of memory coincides with a temporary insanity defense. However, there is disagreement as to whether this should remain the case (Nakic, 2012). Amnesia and dissociative identity disorder have played a role in assessing the

While it may not be true that all theories of mind have distinctive ethical implications (I have trouble seeing any particular ethical implications of epiphenomenalism and occasionalism[2]) we have good reason to take the linkages between these two areas of philosophical study (minds and ethics) seriously. This connection strikes some as so obvious it barely warrants further argument. Daniel Dennett, for example, claims that having a mind automatically grants a level of ethical status to the mind-haver (Dennett, 1996).[3] Humans have minds and rocks don't, therefore humans are beings worthy of moral respect in and of themselves and rocks are not. Strictly speaking, the logic here is spurious, but it is a common sentiment with which most people, including myself, would agree.

The specific beliefs we have about the mind affect much of our ethical thinking. If, for example, mind-equals-brain (and brain only) where the brain goes ethical consideration goes as well. A mind's temporal existence will help determine *when* ethical consideration is warranted (the debate over the possibility of post-mortem harm has to do with this issue). Some formulations of the extended mind thesis view minds as existing in times and places and functioning in ways which require us to reevaluate how we dole out ethical consideration. More specifically, some formulations of the extended mind thesis, if true, would affect how we should respond to bioethical issues such as surrogate and proxy decision-making, and implementation of advanced directives. Though many strong arguments have been made in support of adhering to advance directives and the decisions of an appropriately positioned proxy decision makers,

---

competency of the accused to stand trial as well as being allowable for consideration as mitigating factors in sentencing.

[2] There may well be some form of nihilism which can be drawn from occasionalism if we were to divorce this theory of mind from its religious connotations.

[3] A similar claim was also made by Peter Singer (Singer, 1976)

nearly all these arguments seem vulnerable to the same two kinds of objections.[4]

1- When these directive and decisions are needed the person in question has ceased to exist either because she is either a different (metaphysical) person or no person at all (i.e.- dead or severally demented or permanently unconscious). In cases where she is now a different person she may have very different interests than the person she was before and insofar as she is no longer a person has no interest in her wishes (i.e. – the directive or determination of the proxy) being carried out. This results from a tension between the wishes or goods for the person(s) involved. Rebecca Dresser (1986) states that the accepted legal standard in favor of adhering to advance directives "… rest on the presumption that respect for incompetent patients is best attained by treating them like competent patients." There is a way to read this quotation (out of context) which actually supports Dresser's position. We do in fact best respect incompetent patients by respecting whatever competent wishes they may express. Only in the most extreme cases should we deem a patient to be incompetent in all regards.

Dresser is instead referring to the dichotomy which exists between the incompetent patient as she is now versus the competent person she was in the past. If a patient's incompetence has progressed to such a degree that the patient is no longer numerically/metaphysically contiguous with the drafter of the directive, we would be applying the wishes of one person onto another and failing to respect the wishes or do good for the person who is actually before us at present.

2- We lack evidence that the dictates of the advance directive or the decisions of the

[4] See for example: Dresser, 1986, 1994, 1995, Buchanan and Brock, 1990, Dworkin, 1994, FitzGerald and Wenger, 1997, May, 1997, Tollefsen, 1999, Buford, 2008, and Bonner, 2009.

> proxy are really those of the patient. Directives become outdated and proxies, no matter how close to the patient, still do not know the mind of another person with certainty. Even the best of proxies is subject to bias and emotional interference and as such may not be acting in the best interest of the patient or in accordance with her wishes. (Periyakoil, et al, 2016)

The version of the extended mind thesis I will defend serves to answer these objections in the following way:

1- Though the patient may be permanently unconscious and no longer qualifies as a person in the strictest sense, there exists the possibility that her mind remains a part of this world in a way that morally requires respect and adherence to wishes. If the patient has undergone a radical change of person, it might be the case that her external mind may continue to tie her current self with the self in the past. If indeed the mind supervenes on external (relative to the brain) substrates or social features as well as one's neural anatomy, as the extended mind thesis suggests, we have reason to respect the wishes of the permanently unconscious or radically changed person because of the moral consideration due to all minds.

2- To varying degrees two separate humans can constitute a single agent or person. It may be the case that, say in the case of a long-time married,[5] deeply interdependent couple, the actors are at times separate persons and at other times a single agent. Additionally, the form and kind of information contained in an advance directive might be capable of surviving across various media without a loss of fidelity.

[5] This is not meant to imply marriage is a necessary or sufficient condition for this kind of mental interconnectedness. Such relationships are possible between lifelong friends, the unmarried yet romantically involved couples or blood relatives.

These arguments require rethinking where the mind begins and ends. The traditional (virtually orthodox) view of the mind is that of a closed system directly accessible only from a first-person point of view. Above, answer 1 considers the possibility that the mind's engagement with the world outside the skull may sometimes be more than mere evidence concerning of what a person's mind may have been thinking about. Instead, the extended mind should sometimes be regarded and treated in the same way we regard a contemporary utterance of a person—not evidence of their thought, but literally the thought itself; the only difference being the scaffolding upon which the thought was imprinted.[6] Answer 2 is similar to 1 in that the mind extends past the brain and skull of the individual and may be found, in part, supervening on in neural matter outside her own body, the link forged by historical and social conditions contributing to two people enjoying what Carol Rovane has called a "shared deliberative perspective". In some ways answer 2 may be an easier position to come to accept because at least the physical substrate of the mind is some neural matter or other and it is accessed, from the outside, by the same processes as if the mind were found solely in a single brain or body.

Making the arguments of this dissertation will be a multiple part process. I will first discuss the ongoing problems with advance directives and proxy decision- making in a broad overview. Next, I will explain and defend the extended mind thesis. And lastly, I will discuss the ways in which the extended mind thesis makes us view and respond to the previously mentioned medical ethical issues differently. This will be accomplished over the course of 4 chapters, summarized below:

[6] An utterance is not, strictly speaking, thought itself. An utterance is the result of thought which we treat as the best possible evidence of what was going on in the patient's mind before the utterance was made. This being said we might be able to make the case that vocal utterances are part of the extended mind which imprint on other brains rather than another non-neural media.

Chapter 1

In this section I will outline some of the problems with advance directives and proxy decision- making (Dresser, 1986, Buford, 2008, and Bonner, 2009). The debate over the legitimacy of advance directives and proxy decision-making has been going on for decades and there are many strong arguments which favor respecting such directives and decisions (Buchanan and Brock, 1990, Dworkin, 1994). Problems persist however. There is reason to be concerned that despite the cogency of the arguments raised by Buchanan and Brock et al. some remain unconvinced and have raised similar arguments to those which came before, different perhaps in only small detail (Buford, 2008, Bonner, 2009). Here I will summarize, briefly, some of the current and historical arguments in favor and against advance directives and proxy decision-making and divide them into types dealing with questions of personhood, interpretation and absence (i.e. - a lack of an advance directive or proxy).

Chapter 2

This will be a summary of the arguments for and against the extended mind thesis including my own defense of the position as, at the very least, plausible enough to be taken seriously. This section will include an investigation into exactly what supporters of the extended mind thesis (Clark and Chalmers, Rovane, Levy) are claiming and clearing up some common misconceptions. In short, the extended mind thesis says that if a process taking place outside of a person's brain would be considered a mental process were it taking place in the brain that process is a cognitive process (i.e. – is a process of/part of that person's mind). The note we write to remind ourselves of things are not aids to memory, but rather memories themselves. A pen and paper used to do long division are not crutches for our brains but literal parts of our cognitive machinery. There are, of course, serious limitations to what we can or should consider

to be part of an extended mind. These details will be further fleshed out in this chapter.

An extensive exchange exists between Andy Clark and David Chalmers, and Frederick Adams and Kenneth Aizawa (the principle authors from each side of the debate). I believe I can strengthen Clark and Chalmers's case for the extended mind thesis against the objections of Adams and Aizawa and, by doing so, responsibly use the thesis as a means to strengthen existing arguments for respecting advance directives and proxy decisions as well as to show that their valid application may be more extensive than typically thought.

My method will be to limit the scope of the extended mind. Not all mental states or contents can extend, only those parts not tied to phenomenal experience. By limiting the thesis in this way, I hope to eliminate the confusion between the cognitive and the conscious before it even begins. Insofar as the extended apparatus deals with memory, it will deal with only semantic memory rather than episodic memory, as these terms are defined by Endel Tulving (Tulving, 1983). Semantic memories are those concerned with long-term memories such as use of language, maps of the world, and things generally considered 'common knowledge' but are not drawn from personal experience per se. Episodic memories are those concerned with particular, temporal events in a person's life.[7] To the extent the contents of the extended mind are cognitive they are a subset of semantic memory that I will refer to as 'isomorphic'. Isomorphic in this context refers to the information contained in memories which remains unchanged as it crosses media as opposed to those mental features which may lose fidelity as they move from place to place.[8] I can easily imprint on someone's brain the fact that my birthday is May 7, 1975

[7] For example, I, obviously, learned the words and melody to the song Happy Birthday at some point in my life but this memory is not tied to a particular event or time at which I learned them (semantic). On the other hand, I do remember drilling my few lines (during Mr. Kraft's Algebra II class) for my senior year musical 'Brigadoon' in the spring of 1993 (episodic).

[8] There is an existing term which technically fits the bill here: digital. However, it is sometimes

since it is just a fact, but I will have a much harder time imprinting as completely my happy memories of birthdays past or the anxiety or relief (depending on the context) of entering mid-life.

There is a kind of uneasiness with the extended mind thesis (Adams and Aizawa, 2001, 2005 and 2010, Fodor 2009) these limitations are meant to address. Though no well-studied philosopher of mind will necessarily equate the mind with only those mental features which are conscious, there is a distrust of the extended mind thesis precisely because of the lack of consciousness or even the lack of possibility of consciousness being found in some of the extended scaffolding. The mind does many things in many ways, (conscious/unconscious, intentional/unintentional, voluntary/mechanical/automated) and so long as we limit the extended mind to those features which are independent of consciousness and phenomenal experience we cut certain objections out of the picture before they can even arise.

Chapters 3 and 4

In these chapters I will discuss the implication and effects of the extended mind thesis for advance directives and proxy decision-making. I previously categorized the problems raised with advance directives into three types: problems of personhood, problems of interpretation and problems of absence. Problems of personhood and interpretation will each be handled in separate chapters (3 and 4 respectively). The problems of absence are another matter. Applying the extended mind to these problems is a highly speculative endeavor which would reach far past the limits of this dissertation. We can learn things about a person's mind by examining her environment but applying this reasoning to serious moral decision-making will be a vast

---

difficult to divorce one's self from the technological connotations of this word, thus a new term was chosen.

expansion, or maybe an entirely new understanding of, human psychology. I will include an appendix which outlines my thoughts and summarizes the little work that exists in this area, but I will not be ready to argue for any particular conclusion amidst the speculation stemming from a few interesting studies and observations.

Chapter 3- Problems of Personhood

The question of what makes one person the same person over time is a perennial problem in philosophy. Sameness of memory or continuity of cognitive processes are often criteria cited when philosophers weigh in on this debate. Locke and Reid are the classic examples of this, but others such as Grice (1941), Perry (1975) and Parfit (1984) have made contributions on this issue more recently. The extended mind thesis calls into question what exactly constitutes a memory or cognitive process. Since memory theory of personal identity states that you are (at least) all the persons you remember having been in the past, it stands to reason that the extended mind thesis, and its claims concerning extended memory, will have something to say on the issue of personal identity. If a written advance directive constitutes more than simply evidence of a memory, desire or belief, but rather *is* the memory, desire or belief itself, then is it the case it remains the memory, desire or belief of a patient even if she is not able to contemporaneously express it or expresses a wish which is contrary to the directive. A parallel may be drawn with dispositional memories which are only brought forward under the right circumstances. I have many memories which are not occurrent in my mind at this moment as well as those which I cannot willfully call into my conscious mind, but this does not make them any less my memories.

In the end, I will conclude that a person no matter how demented (short of being permanently unconscious), remains the same person, throughout her biological life, though the

thread connecting her various mental states together can be quite tenuous. This cognitive thread not only includes internal mental phenomena or states, but those which are external. This does not lead, however, to the further conclusion that a directive automatically applies despite being the memory or persistent act of will of that person. Both Ronald Dworkin and Rebecca Dresser are partially correct and partially mistaken in their respective assessment of the legitimacy of advance directives. Contrary to Dwokin, I will argue that even a demented patient can change her mind concerning what he calls critical interests—those beliefs which are concerned with how one wishes her life to go and are about the things most dear to her, and which serve as a lynchpin in his defense of advance directives. For her part, Dresser is incorrect to claim that a demented patient is not the same person as her previously healthy self, but still correct in that we should not, as a matter of inflexible default, obey the dictates of her advance directive. There may be reason to believe the patient has changed her mind in light of changes to the extended elements of her mind.

Chapter 4- Problems of Interpretation

These are problems regarding how to best understand the wishes of the patient. There is always ambiguity in language and this problem is compounded when the drafter of a directive is unable to give further explanation of, or clarification to, her instructions. The strategy here is not to give new tools for interpretation so much as it is to give greater legitimacy to the interpretation of properly socially situated individuals. If a proxy or surrogate has a particular kind of relationship to the patient her interpretation is not a 'best guess' or 'most probable interpretation' of what the patient would want, but rather a revelation of what the patient really did or would desire. The reason for this is that the patient and the proxy can, under certain circumstances, constitute a single mind or agent for limited purposes. If this is the case, the clarity lent by the

proxy is as legitimate as if it were given by the patient herself. The problem still exists of determining who is the most appropriate proxy for the patient, and whether that proxy is "close enough" to share a deliberative perspective.[9] Admittedly, the extended mind will not be as helpful, in this way, in all cases where a proxy or surrogate is needed since people from such close relationships will not always be available. As such, patients will continue to need and use proxies who are not close enough to count as part of their extended minds (or as "sharing agency"). It is even possible that a person will have chosen someone other than her "other self" as a proxy – a situation which I will also discuss in this chapter

For determining the best possible proxy, we will look at the criteria used by Clark, Chalmers and Levy (ready availability, transparency in use, constancy in a person's life and past endorsement) for determining if an external apparatus or separate person qualifies as part of someone's extended mind. The closer a proxy comes to meeting these requirements will be indicative of the extent to which she is a good choice for a proxy.

These same criteria are useful when looking at advance directives. In fact, the nature of an advance directive makes it a perfect candidate for the extended mind as advance directives fit the criteria in ways both practical and moral. If the mind extends, it is possible that the mind of the drafter extends into/onto the directive-making it an observable part of the patient's mind and not just a second-best representation.

[9] The infamous Terri Schiavo case was one concerning who was best to serve as the proxy for Terri, her husband or her parents.

Chapter 1- Advance Directives and Proxy Decisions: Arguments and Short Comings

In this chapter I will discuss the debate concerning the legitimacy and moral force of advance directives and proxy decision making. The history and debate concerning advance directives is a long and winding conversation taking place throughout many disciplines. Many philosophers, legal scholars, medical professionals and editorialists have had much to say on these issues and these opinions and arguments are as nuanced as they are varied. I feel it serves this dissertation well to focus on the most stubborn problems and broadest takes on the issues. The focus of this dissertation is on the implications of a particular theory of mind on a particular subject in ethics, rather than the detailed minutia of individual cases. If I am unable to handle the most general ethical issues raised by advance directives and proxy decision making, it makes little sense to dwell on the details and concerns of more specialized or focused arguments.

To begin, I will start with a handful of the most notable scholars weighing in: Rebecca Dresser (1986, 1994 and1995), Alan Buchanan and Daniel Brock (1990) and Ronald Dworkin, (1994). These works go into useful detail about the problems and benefits of advance directives and proxy decision making but, more importantly for me in the context of this dissertation, they collectively outline several broad categories of perennial concerns which have not been dealt with satisfactorily. The work of others (Bonner, Buford, Nelson, Singer and others) will play roles as well and, as needed, I will refer to them to illustrate finer points as the conversation develops. Though the particular flavor of the arguments changes, many of the same objections were a concern thirty plus years ago are still dogging us today. For example: is the person who drafted the directive the same as the patient in question today? If not, why should we let one person make decisions for another if there was no permission given to do so? Some (Dresser, 1986, Tollefsen, 1999) have gone so far as to call this a form of (or at least draw analogies to)

slavery - a charge as persistent as it is serious. Additionally, with the rate of scientific and medical advancement, is it realistic to think that a directive is valid for more than a few years before becoming hopelessly out of sync with the state of the art in life saving/preserving technology. Our current understanding of minds and persons are inadequate to settle these debates. After identifying and describing a list of persistent problems in this chapter I will have sorted to issues out in such a way that the extended mind thesis (once defended in chapter 2) becomes relevant to the issue and a plausible place to find solutions or at least insight.

Some Basic Distinctions

Advance directives can be divided into two broad categories: the first are instructional advance directives (also called living wills) "whereby a person, when competent, specifies, perhaps only in rather general terms, which types of treatment he or she wishes to have or, more commonly, not have, under certain circumstances, should he or she become incompetent" (Buchanan and Brock, 1990). These kinds of directives "[locate] decision making authority in the formerly competent individual" (Buchanan and Brock, 1990). Directives of the second type are proxy advance directives (or durable power of attorney for health care) in which "a competent individual designates some other individual, or individuals, to serve as the surrogate should the person become incompetent" (Buchanan and Brock, 1990). These directives "enable the formerly competent individual to determine *who*[10] shall decide" (Buchanan and Brock, 1990).

With both kinds of directives the drafter is making a choice about her health care in the event competence is lost. Brock and Buchanan's term "formerly competent individual" is both the source of support for, and reason for resisting the implementation of advance directives.

[10] My emphasis.

Since autonomy is so highly valued (valued above many other important ethical principles by some) we see the importance of respecting the "competent individual." However, the idea of a "formerly competent" patient and perhaps even "former individual" (insofar as 'individual' means 'person') create sticking points. Depending on the degree to which competency is lost, we are faced with different problems as to how we treat the patient, as her interests, and perhaps her very person, have changed. Do we choose as the patient would have decided, were she able? What "stage" of the person are we to consider? Should we consider the interests she had in the past or are we to consider only those interests she has presently or is likely to have in the future?

The two are not as exclusive as it might appear. Even if an instructional directive exists it will still be necessary, from time to time, to find an appropriate proxy to interpret and sort out ambiguities. As such, the instructional directive bears, in practice, some marks of proxy advance directive. Visa versa proxy advance directives (ideally) involve the proxy having been told by the patient what her wishes are generally, bringing instruction into the process. This mixing of the explicit and implicit, and the written and spoken, shows another dimension to the problem. Conflicts and inconsistency can be found if one only looks and considers the evidence more broadly (i.e.- looking for evidence in varied places).

General Arguments for Respecting Advance Directives

In contemporary bioethics, and society in general, great emphasis is put on the individual and the rights of the individual to decide what is best for her and how best to achieve it. This is perhaps more true in health care than other areas of life. We allow a competent person to make medical decisions (even decisions which would be widely considered "bad for her") while she is able and use of advance directives gives her the ability to make decisions for herself in preparation for a time when she will, either permanently or temporarily, not be able to do so.

Aside from society's general proclivity to respect the wishes of the individual other arguments have been put forward in defense of the use of and adherence to advance directives. I will discuss two types (moral arguments and arguments concerning performative utterances), both of which are discussed in Buchanan and Brock (1990).

Though Buchanan and Brock do an exemplary job at summarizing these arguments some still remain unconvinced. Different, or even the same, moral theories can yield contrary results depending on how the reasoner approaches the situation; a problem found throughout moral reasoning. Buchanan and Brock's argument concerning performative utterances is where I will begin my inroad into the extended mind. I will, in this section, briefly outline how I plan to proceed and expand on Brock and Buchanan's arguments from morality and performance.

Moral Arguments

Advance directives are able to do good, morally speaking. This is both a utilitarian claim, which asserts the drafting and adherence to advance directives can lead to a better overall state of affairs, and a deontological claim that it is a good (i.e. - a moral duty) to respect the autonomous choices of a competent individual. If the reader is more inclined to Principlism, as defined by Beauchamp and Childress (1979), advance directives can also act as tools for beneficence, non-maleficence, and justice in addition to respect for autonomy. Here, I will make quick mention of some of the more basic kinds of moral arguments in support of the use and adherence to advance directives according to each of these broad moral theories. There is much more to be said and, likewise, objections to be given. What follows is merely meant to show there are plausible arguments in favor of advance directives within any number of moral world views.

Utilitarian Arguments

The goods of drafting and respecting advance directives benefit both the drafter and those who survive her. The drafter benefits in a few related ways. First, she can protect those interests which she deems most important so her life may unfold as she sees best.[11] The drafter benefits herself by doing what she can to bring it about that at least some of her "surviving interests" are satisfied. If the patient believes she has an interest in not being kept alive on a ventilator if her brain is severely and permanently damaged, she helps ensure the protection of this interest by drafting a directive which makes this interest known. Secondly, though she may never know or experience those interests being protected she benefits from the peace of mind in knowing steps have been taken to ensure her desired future.

Additionally, there are goods beyond those to the directive's drafter. By drafting an advance directive one performs a moral action which benefits those left behind by freeing them not only from emotionally difficult choices, but possible legal ramifications. Survivors are freed from making any number of difficult decisions at a time when they are possibly least able to do so, thus giving them peace of mind as well. Also, we exist in a state of finite health care resources. An advance directive can spell out (explicitly or tacitly) what resources the patient believes should be spent on her care. We do not run the risk of 'wasting' resources on a patient that does not desire them. Of course, there is just as much possibility that the advance directive could request a great many more resources than we might otherwise find appropriate. On balance this is not really a problem. The presence of an advance directive is not a license for the patient to squander scare resources on her own, perhaps hopeless, case; reasonable limits are, and should, be in place as they are now. As concerns resources, advance directives make us, at

[11] More on kinds of interests later in this chapter as we turn to the work of Ronald Dworkin.

worst, no worse off than we otherwise would be and at best we come out some resources ahead. In either case we are never at a net loss of utility.

Deontological Arguments

Advance directives are a way in which a patient extends her autonomy into the future even when she is unable to exercise that autonomy contemporaneously. Though not the only criteria for medical actions, the respect for autonomy carries much weight in medical decision making.[12] The requirement of informed consent, the primacy of a competent patient's contemporary wishes, rights and laws concerning doctor/patient confidentiality and the right to refuse even lifesaving treatment, all stem from our acknowledgement of the importance of a respect for autonomy.

Additionally, the presence and use of advance directives creates a useful and clear rule for us to follow and a duty to adhere to. It is an open and problematic question as to when a patient becomes incompetent, though there are obvious cases at the extremes (permanently unconscious patients or the dead for example). An advance directive spells out the conditions under which a patient would no longer consider herself, herself. The decision as to when the patient's life, as she desired it, has ended is no longer a medical decision per se, but instead a choice exercised by the individual based on her preference for what kind of a life is worth living. Incompetence becomes a moral choice of the individual and not a biological or medical

[12] This, however, is not the only principle at work and the respect for autonomy is not absolute. Principlism (discussed in the next section) argues that patient autonomy can be overridden under the appropriate circumstances. But suffice it to say, generally, the respect for autonomy is very important in moral decision making.

determination, thus helping to sweep away these difficult issues of determining when competence is lost.[13,14]

It is not an unreasonable understanding of an advance directive to say that it is a set of instructions by the patient to express to others her presumptively authoritative desires concerning how she should be treated when she appears, by all reasonable measures, to have lost competency (i.e.- we determine she has lost competency) and not a requirement that competency has actually been lost.[15]

Why These Moral Arguments Aren't Enough

Should utility or rights and duties rule the day, morally speaking? Though this is not the oldest debate in ethics this distinction is a clear line of demarcation in modern ethical discourse. In an abstracted bioethical scenario even a first-semester philosophy student can see the logic of

[13] Mutatis mutandis, a similar argument in support of this point can be extracted from an exchange between Cole (1993) and Tomlinson (1993) regarding irreversibility as a criterion for a determination of death. Cole argues that irreversibility is a criterion for the concept of death to which Tomlinson replies that irreversibility is [at most] part of the determination of death, not actual part of what 'death' means in an ordinary sense.

[14] A case could arise where a patient is technically competent (perhaps suffering from total locked in syndrome) but from the 'outside' all indications are that the patient has lost competency (i.e. – is permanently unconscious). There are some accounts of competency which explicitly require communicate of desire or wishes (Applebaum and Grisso, 1989). This seems to be hair splitting since it is entirely possible for a person to understand her situation but unable to tell anyone she understands. For practical reasons, we would nonetheless need to treat this patient as incompetent. If this apparent loss of competency conforms to the circumstance spelled out in the advance directive we are justified in carrying out those instructions despite the patient being technically competent since, from our perspective, the patient is in a state which conforms to her previous moral choice as being one of incompetence.

[15] I have heard this situation, a conscious yet immobile patient, described as a neurologist worst nightmare. And undoubtedly mistaken determinations of competency have been made (see the cases of Jahi McMath and Martin Pistorius for examples) just as mistakes have been and will be made in the determination of biological death. However, until such time brain science advances in such a way as to avoid these errors by opening up the minds of such immobile patients the possibility of a misdiagnosis is, quite unfortunately, just part of the hand we are dealt.

both the utilitarian and deontological arguments for action. But, outside of a bioethics classroom, decisions are reality, not abstractions. It is highly likely that a family dealing with a dying loved one won't even have the minimal formal training of our philosophical novice, let alone the medical, ethical or logical chops to fully appreciate the gravity of the situation their loved one faces.

Just as we can see the merits and follow the arguments of each side, we can also see that the utilitarian and deontologists disagree internally. Moral arguments have been given as to why we should adhere to advance directives, but what about the moral reasons to *not* do so? Briefly:

Utilitarian

The patient is unconscious or, at best, minimally aware of what is happening to her. There is little, if any, good left in her life. However, the family will still need to be considered after the patient has died. The family would be happier if the patient were kept alive, contrary to the instructions in the advance directive. We create an overall "happier" world by keeping the patient alive, even if only for a short time.

Deontological

Directives vary, and a rule with variance is no rule at all. The maxim by which the patient determined what should be done is faulty, or she was not autonomous (fully informed) about what her future situation would be like. In the absence of relevant information, we cannot say that the patient has made an autonomous choice, therefore we are at liberty to disregard the directive. Though we could apply this same reasoning to any decision made by an agent (e.g. - we never have *truly* complete autonomy when making choices), cases involving advance directives are importantly different in that, by the time the directive comes into play, it is often impossible, even in principle, to get more information to the person making the choice.

Perhaps you do not find these brief sketches of arguments convincing, but others might see strength in these babies of ideas. Competing goods and conflicting rules are a recipe for disagreement and a poor forum for finding answers to difficult questions.

Beauchamp and Childress (1979) saw that even though utilitarian and deontologists disagreed about which theory should be used to answer ethical questions, theorists agree, almost universally, as to what was ultimately important when it came to actually making moral decisions. They identified four principles (respect for autonomy, beneficence, non-maleficence and justice) that all parties believe are important, worth protecting and upholding, and thus creating a new theory: 'Principlism'. So, maybe this is the moral tack we should take.

Principlist Arguments

The adherence to advance directives upholds, in various ways, the moral principles we all believe are important in making moral choices.

- *Respect for Autonomy* - As was previous mentioned, the respect for advance directives allows a patient to decide for herself how her life should go. If we do not respect the instructions in a directive we are saying to the patient that her wishes are secondary to the desires of other, perhaps less interested, parties.

- *Beneficence and Non-maleficence*[16] - The patient can decide and make known what she believes to be in her best interest or harmful to her. This is above and beyond the respect for autonomy these goods are most often associated with. As with the deontological

[16] Though identified as separate principles by Beauchamp and Childress in such a short summary handling them together is efficient as they about the closely related concerns of harms and benefits.

argument about competency being a moral choice, so too are the benefits and burdens borne by the patient.

- *Justice* - Medical resources are sometimes scarce. If the patient refuses treatment through her advance directive, should the circumstances spelled out in her advance directive obtain, we are perhaps acting unjustly toward other patients if we continue to use resources which are verboten according to the directive. To follow the instructions in an advance directive not only respects the patient's wishes and, thus her autonomy, but also can forward the cause of justice in medicine more generally by freeing up resources.

Again, we could just as easily formulate our arguments to come to the opposite conclusion (which would be the right approach given the right circumstance) and are forced to make choices that the principlist position is not equipped to make. We also face the general problem of weight within principlism. How do we weigh one interest over another? When does justice outweigh a respect for autonomy? How much harm can we do in the service of helping the patient?

Even if all moral theories could agree on which principles are most important to uphold, we still would not be able to settle the question as to the status of advance directives through moral reasoning alone. Though theorists can see the value in the other side's arguments, in the end utilitarians like the utilitarian approach and deontologist like deontology. When faced with conflicting conclusions from one, or more, moral theories we cannot apply to those moral theories in order to resolve the debate. I find this akin to using a pen to write on itself. Something about the claim seems plausible, but in practice we just can't seem to make it work.

To achieve some sort of a definitive result in a moral debate either we will have to argue for one position or the other by way of a metaethical theory (which only pushes the problem off to another level of analysis where more conflicts are likely to arise) or apply to a non-moral set of arguments. The extended mind thesis is one such set of non-moral arguments. If we have good reason to view the mind in some way apart for the orthodoxy some of these uncertainties may fall away.

Advance Directives as Performative Utterances

Buchanan and Brock make further arguments in support of advance directives by means of viewing the directives as performative acts. In doing so there are two primary ways of viewing advance directives: as guidance or mere evidence. To view an advance directive as guidance is to see the directive as instructions given by the drafter as to how to proceed. If viewed as evidence, advance directives are little more than an aid in determining how the patient would decide rather than what she in fact decided. Buchanan and Brock argue the correct view is that advance directives are performative acts that constitute acts of will (Buchanan and Brock, 1990) and are therefore not merely evidence of the wishes of the patient but *are* the wishes of the patient. We do not read a last will and testament and use this as evidence of how a person would want us to distribute her estate but rather we understand it as a set of instructions as to actually how to distribute her estate.[17] If a terminally ill patient is conscious and informed we (as a general rule) will respect her decision to refuse treatment, even lifesaving treatment. We do not treat the contemporaneous choice as mere evidence of what the patient desires. The

[17] That is, of course, unless there is some ambiguity present. This is a potential problem with advance directives as well, which will be dealt with throughout this dissertation.

contemporaneous utterance of the patient is a set of instructions the patient intends to be carried out.

Buchanan and Brock are correct in their assessment and the extended mind thesis will help further argue this point. If we consider the execution of a last will and testament we take the document at its word and treat it as the actual expressed wishes of the deceased, not as guidance or suggestion. When a will is executed there is no generally considered question as to the existential status of the drafter; she is dead and therefore not only no longer the person who drafted the will, but is in truth no person at all. Yet we follow the text of the will as if she was right there in the room telling us herself how her worldly possession should be disposed. The case of an incompetent patient with a standing advance directive is not that dissimilar. The patient is unable to express her wishes at present (here due to infirmity rather than death) yet we call into question the validity of those words. It could be argued that we have, on the face of things, a *stronger* obligation to follow the instructions of an advanced directive versus a last will and testament as the existential status of the patient is not clear. Dead is dead and therefore no person exists, but infirmity does not imply non-existence. Arguments can, and have been made (and will be throughout this dissertation) that the incompetent patient is numerically and metaphysically identical to the drafter of the directive. If this is the case the patient is, in a manner of speaking, in the room telling us herself how to handle her current medical predicament.

The extended mind thesis makes an argument in favor of the sameness of persons in situations where an advance directive is in question. If the mind can extend it is possible that a patient can extend her will into the future through the use of external (relative to her brain)

apparatuses. The mind may not entirely disappear even after the biological death of the patient. This being the case, there remain patient interests and wishes to consider.

Performative actions are a foot in the door to making the extended mind thesis an important consideration in manners involving advance directives and proxy decision making. The forthcoming argument will be that the directive is a persistent act of will of the patient. This being the case, the person exists as an entity worthy of moral consideration so long as the directive is applicable. Directives are not issues of absent persons, but rather, a person whose existence is different than we traditionally consider. The orthodox view is that whatever it is that constitutes a person is so tied with the biological body/brain that where the body goes, so goes the person. The extended mind means this may not always be the case. In situations where a directive needs to be enacted either the person is constituted by a combination of the biological being and written directive (the directive being perhaps the only accessible mental state of the patient) or solely constituted by the directive (an externalized, and continuing act of will).

Problems with Advance Directives

The above are reasons for respecting advance directives but what is missing so far are the arguments that adherence to advance directives is a moral imperative. Although reliance on either instructional or proxy directives possess unique problems, Buchanan and Brock point out four problems common to both:

1- A vast majority of people have not issued either such directive. A 2014 study published in The American Journal of Preventive Medicine found that:

> Of the 7946 respondents, 26.3% had an advance directive. The most frequently reported reason for not having one was lack of awareness. Advance directive completion was associated with older age, more education, and higher income and was less frequent among non-white respondents. Respondents with advance directives also were more likely to report having a chronic disease and a regular source of care. Advance directives were less frequent among those who reported not knowing if they had an EOL [end of life] concern.

Little more than 1 in 4 Americans have completed advance directives. A minority to be sure but, certainly not a trivial number of people. However, disparities exist in terms of who completes the directives and under what circumstances. Advance directives are not distributed widely or evenly enough to be as socially impactful as we could hope.

2- Some individuals were never competent to issue such directives (such as the young or those born with cognitive disabilities).

3- When an advance directive does exist, it may still be, necessary to find an authority who can interpret it and to allow medical personnel to implement the directive most appropriately. When an advance directive is ambiguous or vague, reasonable people may disagree as to the proper interpretation of the directive. The exact reasons for the ambiguity and uncertainty are many but all center around a lack of knowledge about the patient's wishes and her inability to anticipate all future circumstance regarding her health and the state of medical science.

4- Patients who are incompetent to a severe enough degree may no longer be persons and thus, to the extent they have interests, they are interests which are very different from a patient possessed of personhood. This is a different kind of problem all together. Issues of surviving interest have been discussed widely in the philosophical literature, primarily as they relate to the dead. Rather than weigh in on the issues of the rights or potential agency of the dead, I will limit this part of the dissertation to patients who are alive (either self-sustained or via medical assistance) who need to have choices made for them.[18]

[18] There exists a fifth type of problem in that health care providers have been known to ignore or evade advance directives (Connors, et al.). This is a serious problem to be sure, but one that does not need addressing, per se, within this dissertation. If, as I will argue, an advance directive has just as much moral force as a contemporaneously uttered desire of a patient, the moral imperative to pay attention to advance directives would be just as strong as those involved with respecting

These four broad categories of problems can be further subdivided or otherwise parsed out. The following sections will elaborate on these basic concerns and introduce a few additional ones. The work of Rebecca Dresser will examine the limitations of our knowledge as it relates to advance directives and Ronald Dworkin will consider the roles our different types of interests play in making medical choices for others.

Rebecca Dresser

In cases where an advance directive exists we may not be sure if it is a legitimate directive or, if legitimate, there may be questions as to exactly what the drafter intended. Reasons for questioning the legitimacy of a directive or difficulty in interpreting it are varied and distinct problems. Rebecca Dresser in her 1986 article *Life, Death, and Incompetent Patients: Conceptual Infirmities and Hidden Values in the Law* summarizes the main problems we face when considering advance directives. These problems are still dogging us today as they are also discussed in several places in more recent literature (Buford, 2008 Bonner, 2009 and FitzGerald and Wenger, 1997). I will consider these problems in four broad categories.

1- *Change of person* (Dresser, 1986, Buford, 2008)

The current patient and the directive drafter are so different psychologically that we should consider them different persons. Since it is wrong for one person to make decisions for another (without consent) it would also be wrong to allow a different past person to make decisions for a present person. This problem is serious enough to some that they consider this may be a form of tyranny or slavery (Dresser, 1986, Tollefsen, 1999).

A change of person can come in two forms: a process resulting in a different person

---

the wishes of a patient.

associated with a persisting body, as for example, might happen via dementia (pleasantly or otherwise) and a complete loss of personhood. If a person is demented to a severe enough degree we may have reason to say, under certain neo-Lockean understandings of personal identity, that the patient is a different person than the person who drafted the directive. Though dementia is more often than not considered and experienced as a harm for the patient there are rare cases of a patient, who is "pleasantly demented," meaning a patient who is diagnosed with some form of dementia but whose life is not burdensome, though perhaps intellectually simpler than it had been prior to illness. Regardless of the case, if the patient and the drafter are metaphysically distinct persons it may be wrong to apply the directive.

In those cases where the patient is no longer a person (permanently unconscious or impaired to the point of complete unresponsiveness) it is harder to see that the slavery argument applies. For the "master/slave" relationship to exist there needs to be a master and a slave. The drafter of the directive is the analogous master in this case, but since the patient subject to the directive fails to meet the criteria for personhood, there is no one to be the slave. The 'master' or drafter of the directive has made a choice for himself in the future but there is no person at that future time, therefore there is no master/slave relationship.

Even if we dismiss the slavery argument here there are still other issue we must contend with, such as why should we apply the directive in the first place, as there is no person to whom the directive was meant to apply? If the patient is no longer a person does it matter at all to the patient, morally, how we treat her? Do the permanently unconscious have interests which need to be protected?

2- *Out of date directives* (Dresser, 1986, Bonner, 2009)

No one can predict with certainty what medical advances and new treatments will be available at times after a directive is issued. President Franklin Roosevelt pleaded for an all-out attack on polio in 1944 and within a decade a vaccine was available which had an 80-90% success rate at preventing paralytic polio (historyofvaccines.org).[19] Cancer, on the other hand, has proven far more stubborn.[20] Sometimes a treatment for a fatal or debilitating condition is just off the horizon and others no cure or therapy is in sight. If the patient had this information at the time the patient may have constructed her directive differently.

Consider the following example and dilemma faced by patients today. At this time there is no cure for Amyotrophic Lateral Sclerosis (ALS or "Lou Gehrig's Disease"), a neurodegenerative disease characterized by a progressive loss of voluntary and involuntary muscle control. Getting such a diagnosis is a death sentence, most likely by respiratory failure, which will come on average in about five years (als.org).[21] If a patient is given such a diagnosis it is not unreasonable for them to draft an advance directive which states that no extraordinary measures should be taken to save her life should she develop some other life-threatening condition which "beat ALS to the punch," so to speak. However, if a cure for ALS were on the horizon this might change the patient's mind about her previous choice. If getting cured of deadly pneumonia will help her live long enough to see a cure of ALS she might well decide to

[19] The initial investigations into a possible immunity to polio began in 1910 but once a concerted effort was begun the field advanced very rapidly.

[20] President Nixon declared a "war on cancer" in 1971. Forty plus years on, treatments have been developed and outcomes improved in some cases, but our efforts have mostly taught us that cancer is a far more insidious and complex disease than we initially assumed (Mukherjee, 2010).

[21] There are cases of patients dramatically beating this average. Notably the physicist Steven Hawking has been living with ALS since 1963. Meaning, as of this writing, he has lived with ALS for more than fifty years. Cases such as this are extraordinarily rare.

undergo the treatment for the acute condition.

3- *Inability to cover all contingencies* (Dresser, 1986, Bonner, 2009, FitzGerald and Wenger, 1997) Advance directives are finite documents meant to protect a patient's interests and wishes in a world where a seemingly infinite number of things are trying to thwart those interests and wishes. No directive can account for every possibility, not even the known possibilities. If the patient is too specific the patient runs the risk of not having what would be her wishes carried out in certain circumstances. If the patient tries to get around this issue by being "usefully vague" it is possible life-saving or sustaining treatment will be withheld in circumstances not objectionable to the patient.

If I draft a directive which states I should not be kept alive on any kind of heart lung machine should I suffer traumatic brain injury as the result of an *illness or infection* how are we to understand this instruction if I become seriously brain damaged as the result of an *accident*? One reading here is that what is important is the traumatic brain injury itself and not its cause, but who is to say? It is not implausible for a person to view an accident as somehow "unnatural" and thus consent to treatment while an illness needs to be allowed to run its course because it is "natural."

Usefully vague use of terms, such as "heroic" (as in 'do not take heroic steps to save my life'), are common in advance directives, and although this term is easily enough understood in polite conversation, it is legally and morally difficult to nail down. Perhaps I consider respiratory intubation a heroic action, whereas the medical community sees it as a serious, yet routine step in trauma care.[22] Things could go the other way just as easily. I see an extended period on life support in intensive care as routine, but my caregivers consider it "heroic" given

[22] Thanks to Dr. Jeni Shreve Balawender for helping me with this example.

the extent of my injuries.

4- *Change of Heart* (Dresser, 1986)

Even if personhood remains constant through illness or injury it is possible a patient may change her mind concerning what treatments or conditions are and are not acceptable and her directive does not reflect her most current feelings or wishes. This change of heart could result from a number of causes. Something as stark as a religious conversion could be the impetuous for some, while for others it could be the result of contemplation and considered opinion as the patient progresses through her life.

Additionally, what people value and what people deem a life worth living can change as the circumstances of a person's life unfold and new circumstances arise. I once thought I wouldn't want to live a life where I was dependent on a wheelchair. Today I would be more willing to live with a severe physical disability than I would have in the past so long as my mind remained reasonably sharp. I put more value today on my mind than my ever-deteriorating middle-aged body.

A person's change of heart might also be conditional. The patient's overall values may not have changed but circumstance gives them a desire for a little wiggle room. A patient might still think that living with the aid of machines is not a long-term prospect she relishes, but might be willing to suffer what she thinks is an indignity for a short while if it means getting to meet her first grandchild.

An advance directive being subject to possible changes of heart is not the same thing as being subject to the moment to moment whims of the patient. If it can be shown that the patient's wishes have been changed under genuine, or at least not spurious, consideration, we should take seriously the possibility that the directive and the wishes of the patient have

diverged.

As stated, a common thread runs through all these objections: we lack certainty in our knowledge of what the patient would or does want. Accordingly, I will collectively refer to these four objections as KPADS (Knowledge Problems for Advance Directives). KPADS show there is a disconnect present between the reason for, and problems with, adhering to advance directives. If we follow the directive we are considering what were (at least at one time, to the best of our knowledge) the wishes of the patient. But in doing do we also run the risk of sticking to the strict letter of the directive while ignoring its semantic intent or the current views of the patient. Buchanan and Brock draw moral and linguistic conclusions to persuade us that we should follow advance directives, while Dresser (et al) have expressed concerns about our knowledge of the true wishes of the incompetent patient. This being the case, we can believe Buchanan and Brock wholeheartedly yet still express concerns we might be acting contrary to the wishes of the patient.

If indeed some of the most pressing arguments against the use and respect of advance directives are KPADs, an adequate theory of mind may be able to alleviate at least some of these concerns. Most theories of mind assume the mind to be the kind of thing that is accessible to a single human or person and only from a first-person point of view. Additionally, some theories of mind say the mind is what constitutes a single person and vice versa. If this is the case, then for any mind there is one, and only one, person who has privileged access to a given mind or is constituted by that mind and therefore only one person who can be certain of the contents of that mind. However, there is another possibility being discussed in the philosophy of mind literature: the extended mind thesis.

In short, the extended mind thesis states that the mind extends past the skin and skull and

is, in part, constituted by the extra-neural, observable world. If this is the case the mind is no longer an entirely closed entity observable only from a first-person perspective. Other persons, if socially situated in the right way, can have *direct* access (i.e.- not inferential or inductive based on external behavior) to the contents of another mind either through empirical observation of those contents, or by sharing in the contents of that mind (i.e.- one mind spanning two persons or brains).

If the extended mind thesis is true, the possibility exists that the limitations of our knowledge of the wishes of the incompetent patient are not so intractable or insurmountable. Written advance directives, on this view, under certain circumstances, are not merely evidence of a person's wishes or, but literally the observable content of that patient's mind (the directive is the patient's act of will) which can and should be treated as authoritative in the same way a patient's competent utterances would be. We are faced with a seeming disanalogy in that a competent (or at least conscious) patient can be asked follow-up questions or give further clarification. This should not concern us however. The competent patient may not need to or wish to give further clarification. If this is the case the last stated wishes should be considered be the only state of the person's mind at issue. An advance directive is similar to a patient in this case. The directive does not change. It is a static mental state as is the mental state of the competent who does not change his mind or does not make a change of his mind known. There still exists the possibility of a conflict between what the words on the paper say and the proclamations of a properly situated individual; the directive says X but the loved one says Y. This conflict is nothing new, however, the extended mind thesis allows us to view the situation differently in that the mind of the patient is not absent while this conflict is sorted through, but rather is located in a different place then orthodoxy have us believe. This placement of the mind

makes it accessible to people other than the one whose mind it is. This may allow us to resolve the conflict because we have additional tools at our disposal for discerning the wishes of the patient.

In addition, many, if not most, advance directives are drafted after consultation and with the help of others. A husband and wife, for example, may spend a great deal of time talking over the details of the directive. The extended mind thesis opens the door to the possibility that the husband and wife are not two separate minds consulting on a problem but rather a single mind working at a task at hand. If this is the case the 'proxy' is not a proxy at all but *rather* a revealer of the decisions made by the mind constituted by two human brains.

It is unlikely that any method of implementation and interpretation of advance directives will always yield results beyond reproach and above controversy. Even relying on the seemingly competent testimony of a conscious patient is no guarantee that we will act in the best interest of the patient or that the patient is even fully aware of what her best interests are. But, if the extended mind thesis is true we will have done much to increase the possibility of an acceptable level of confidence, since, among other things, the mind is no longer a closed, private, first person only object or system.

Ronald Dworkin

Ronald Dworkin argues in favor of adherence to advance directives in cases were the patient is no longer competent to make health care decisions. Previous arguments which support his position have centered on the idea that a person is the best judge of what is in their best interest and therefore we should allow the individual to decide. He calls this the 'evidentiary view' of autonomy. The stated desires of a competent individual are the best evidence of what the person thinks is in her best interest. Dworkin disagrees with this view stating: "People are

not the best judges of what would be in their own best interest would be under circumstances they have never encountered and in which their preferences and desires may drastically have changed" (Dworkin, 1994). I may believe I know what I would desire in cases of medical catastrophe, but once the situation is at hand I really do not know how I will react or how the circumstances may change my desires.

Dworkin asks us to consider the following example: a Jehovah Witness has been in a serious accident and requires a blood transfusion. The patient has long believed it is wrong to receive a blood transfusion (as Witness doctrine teaches) and if asked at any time before the accident if she would want a transfusion should the need occur she would say 'no'. However once in the hospital after the accident and confronted with what may be a life and death choice, she may allow the transfusion in an effort to save her mortal life. These conflicting values are not a sign of incompetence, but rather a change of heart or mind once novel circumstances are actually encountered

Not only do people change their minds when circumstances dictate, but people also make autonomous choices contrary to their best interest. The choice to smoke, for example, is hard to see as a choice one makes to improve her life or one that is ultimately in her best interest. Yet hundreds of millions of people worldwide make this choice with knowledge of the deleterious health effects this habit has. We can also view self-sacrifice as an autonomous choice which is counter to the agent's best interests, yet we applaud such choices in many cases. These sorts of cases lead to a serious difficulty with the evidentiary view, as illustrated by this argument:

P1- People make the autonomous choice to smoke.

P2- This choice is harmful and thus counter to the person's best interest.

P3- We respect this choice as it was the autonomous choice of the person.

P4- We respect a choice which is harmful and against the best interest of the person.

Therefore, respecting autonomy cannot be (just) about protecting the welfare of individuals.

Given this difficulty with letting people make a decision counter to their welfare or best interests, Dworkin suggests instead we consider an 'integrity view' of autonomy. This view states: "autonomy … derives from the capacity it protects: the capacity to express one's own character—values, commitments, convictions, and critical as well as experiential interests—in the life one leads" (Dworkin, 1994).

By accepting the integrity view of autonomy we have not claimed that a person will not make choice against their best interests. People can and do make 'bad' choices and even choices which are inconsistent with one another. What the integrity views says is that persons have the capacity and the right to make these choices. The capacity for autonomy and the right to express that autonomy gives each person the chance to determine the direction of her life rather than be "led along" (Dworkin, 1994).[23]

The advance directive of a competent person is a means by which she shapes the course of her life. But once the person becomes an incompetent patient we need to ask ourselves from where does an advance directive derive its authority. If there are doubts about the sameness of the person who drafted the directive and the now incompetent patient before us, do we have

[23] This is the difference between a substantive versus procedural view of autonomy. If we held a substantive view, a view that says 'autonomy is this and only this' any choices made by the agent would either conform to the substantive view and thus be a competent, autonomous choice or be an incompetent choice by virtue of not conforming. The procedural view, by contrast, says that as long as the decision was not A, B or C (coerced, made in ignorance etc.) the choice is competent and autonomous even if it is not objectively in the patient's best interest. (Varelius, 2006)

reason to ignore the directive or does it stand? What happens if we assume the sameness of person? Should we enforce the dictates of the directive even if the incompetent patient protests?

To answer these questions Dworkin invokes the idea of precedent autonomy. This autonomy refers to the competent wishes of the person in the past regarding the current state, that of an incompetent individual, which she now faces. The integrity view of autonomy gives us strong reason to respect the advance directive of the patient. The directive is a statement of the wishes of the autonomous person as to how she wishes her life to go. It is a statement of the upmost values of the person: the circumstance under which she believes her life may not be worth living. Perhaps she is wrong and in truth her present state as an incompetent patient is quite pleasant and well worth living -what then? As Dworkin argued, the right to autonomy cannot be inextricably tied into a person's best interest since we can, do and should respect autonomous choices of a person which are counter to best interests of that person. In order to disassociate the autonomy from the best interest of the patient Dworkin considers the fact that not all interests are same.

Experiential versus Critical Interests[24]

Although Dworkin argues the right to autonomy is separate from whether or not a person makes choices consistent with her best interests, he does not think interests are wholly unimportant. All persons have interests and our choices are the means by which we protect those interests. One means by which we make choices to protect our interests are advance directives. But the question arises: what kinds of interests? I have an interest to not be in pain but I also have an interest in living as long as I can do so in reasonably good health. What if a painful surgery is the means by which I can protect my interest in living a longer, healthier life?

[24] This distinction and the subsequent definitions come from (Dworkin, 1994).

Interests sometimes conflict and when they do, it is not always obvious what interests should win out.

Knowledge problems aside, there is one other important kind of argument given in support of following advance directives. Rather than focus on the continuity of person, Dworkin (Dworkin, 1994) chooses to consider the interest the demented patient currently has contrasted with those interests she had in the past. He divides interests into two kinds, experiential and critical.

Experiential interests are those involved, at the most basic level, with the interest to experience pleasure and not experience pain. What is of experiential interest to one person may be of no interest to another. I happen to love watching Star Trek but my sister would rather do just about anything else. This is not a matter of one of us being right and the other wrong. It is a matter of what experiences are preferred by the individual. Critical interests are those interests one has concerning how one's life should go or what makes a life worthwhile or in developing familial relationships while others will more highly value friendship. What is the same is that a person will think her life is better or worse off should the interests be met or not.

The important distinction between these two kinds of interests is the degree to which a person need be aware of, or present for, them in order for them to be fulfilled. Experiential interests require awareness of those interests by the person to whom they belong. If there is no experiencer there is no experiential interest. A deaf person does not have an experiential interest in not being exposed to annoyingly loud sounds[25] nor does a corpse have an experiential interest in not being caused pain or in having a good meal.

[25] This is assuming the only problem with the sound is that it would be annoyingly loud should someone be able to hear it. Sounds which are at the right frequency and loud enough cannot only cause damage to the human body but they could actually kill you, whether you hear them or

Critical interests on the other hand do not require (but also do not exclude) the presence or awareness of the interest bearer. Because critical interests do not depend on the subject's awareness, Dworkin believes critical interests of a person may survive the person as a biological organism. Say for instance I, as a competent person, develop a critical interest in seeing my wife succeed in her job. Later in life I become demented and am unaware of my previous critical interest in my wife's work. If it is the case I can only have interests I am aware of, it seems this interest has disappeared with that aspect of my personhood. But, because awareness of an interest is not a requirement for having an interest my loss of awareness is not grounds for claiming the interest is lost. Even though I am not aware I have a critical interest in seeing my wife succeed at work I have an interest in this nonetheless.

If we actually believed we only had interests of which we were aware, or could in principle be aware, the debate about advance directives would look very differently. Say I draft a directive which gives specific instructions about how my medical treatment should go if I were to end up in a persistent vegetative state. While in a vegetative state, by definition, I have no experiential interests. Since I am not capable of experiencing anything, therefore the only interests which are left are critical interests. However, if we limit interests to those things of which we are aware I have no critical interests while in a vegetative state either. If this is the case, my advance directive does not apply, at least as far as the directive is a document concerned with one's interests; I am not aware of my interests, therefore I do not have any interests. This may strike the reader as absurd, as well it should. Advance directives are based on the assumption that we have interests of which we are not, or in the future may not be, aware

---

not. Eardrums rupture at about 150 dB (the sound one experience when standing next to a jet engine during takeoff), and death can occur anywhere between 180-200 dB (Anthony, 2014).

of. Some advance directives are drafted out of concern for the patient's experimental interests. If I do not wish to live out my final days in pain, I may draft a directive which makes this desire known. However, I argue, the very nature of advance directives and the circumstances under which they are to be implemented are first and foremost aimed at protecting critical interests. If a patient has a directive aimed at ensuring a minimization of pain, even if it means that patient's death, the directive only comes into play once the patient is incompetent or incapable of making decisions. So long as the patient remains competent and communicative she is free to decide whether or not her levels of pain are tolerable and/or worth living through even though her directive may contradict her current expressed wishes. So long as the patient is able to make competent choices no advance directive comes into play and no proxy decision maker is needed. Directives only come into play when the patient has lost either capacity for rational decision making (PVS etc.) or has lost this capacity to a very large degree (the severely demented). In such cases patients do not have experiential interests or, at most, have extremely limited experiential interests (like those associated with physical pain and discomfort) so there is only their critical interest to consider as concerns the advance directive. But given that we have not eliminated experiential interests in all cases we are, therefore, still faced with the problem of weighing one type of interest against another.

Dworkin argues that the respect we show for advance directives should be motivated by a respect for autonomy which is best done by considering the critical interests of the patient. Suppose a patient writes an advance directive which states that she does not want any lifesaving treatment to be given should her dementia advance to a certain degree. She believes living a life of dependency and incompetence is not worth living and counter to the life she has lived to that point. She fears the embarrassment that would come along with living in such state. Here there

is an expression of both kinds of interests. On the one hand, the patient's desire to protect her critical interests in living the kind of life she desires, one of independence and freedom. On the other hand, she wishes to avoid embarrassment, thus expressing an experiential interest. Though we can all sympathize with the desire to avoid embarrassment, there is an argument to be made that this itself is not a reason for respecting an advance directive. A patient can become so cognitively impaired that she lacks the ability to feel embarrassment. If we concerned ourselves with preventing the patient's embarrassment we succeed whether we adhere to the directive or not, since there is no feeling embarrassment in her present state. If we consider her critical interests, we do have reason for adhering to the advance directive over disregarding it. As Dworkin, argues critical interests can survive the person. In fact, some argue (Nagel, 1970, Feinberg, 1984,) critical interests continue on even after the biological death of a person. Talk of surviving interests seems to be tricky business. Our common use of language is such that in order for there to be an interest there has to be one whose interest it is; just as there cannot be swimming without a swimmer. However, there is no serious controversy in the execution of a last will and testament, which is, after all, the expressed interest of a person who is no longer living. The extended mind does not change this linguistic difficulty, but instead allows for the possibility that a person is not actually non-existent when they are dead. Insofar as the mind and a person are tied together in some way, we might be able to make the case that so long as the extended parts of cognition remain something of that person remains.

Many theories of personal identity argue that what constitutes a person, in the philosophical rather than biological sense, is not the living body. As far back as the pre-Socratics (notably Anaxagoras and Empedocles) it has been argued that a 'person' is not the same thing as her body. Plato (*Alcibiades* and *Phaedo*) and Aristotle (*On the Soul*) wrote on the

subject. Descartes gave us the, now rightfully maligned, modern conception of substance dualism in which mind and body were distinct substances. Locke argues for the difference between a person and a human animal.[26] Today use of the word 'person' in a philosophical setting often denotes a set of psychological characteristics which, depending on who exactly you ask, include consciousness, self-consciousness, sentience, future retardancy, agency, autonomy, empathy, desires, acts of will, emotions, memories, dispositions, introspection and a host of others. Not all philosophers separate the body from the person so starkly. Thomas Reid, in responding to Locke, argued for the necessity of a body for personhood to continue unbroken through time. He agreed with Locke that mental characteristics were important, but not the whole story.

Insofar as a person is her mind, the extended mind thesis makes it possible that some aspects of a person's mind persist after the person's biological death and does so in a way that is not the metaphysical maze of substance dualism or our more contemporary conversation about surviving interests. (the debate over the possibility and plausibility of surviving interests was a lively one in the mid to late 1980s. For more see: Callahan, 1987, Feinberg, 1984, Marquis, 1985, Nagel, 1979, Partridge, 1981, Pitcher, 1984, Wicclair and DeVita, 2004).

If a memory can be encoded in non-neural hardware (such as a notebook) it may still be a memory after the neural wetware from whence it came has ceased functioning. If an act of will can be expressed on paper it may remain an act of will after the drafter has died. If I am my mind, and my mind can extend, I can extend my existence into the future past my biological death (at least for certain purposes).

[26] Locke most often used the terms 'person' and 'man'. 'Human animal' captures his sentiment and I believe more adequately describes what Locke was after according to our modern ears.

None of this is to say critical interests are, a priori, more important than experiential interests within the context of one's life. I have a critical interest in going on living, working and generally trying to be a productive member of society. But if I were undergoing torture at the hands of a cruel prison master it may be the case that my experiential interests of no longer undergoing inhumane treatment, even if it means my death, might be more important or pressing than my interests in going on living. The opposite can also be true. I have an interest in not suffering severe pain but if doing so, due to, say a lengthy recovery from surgery, can improve my quality of life or substantially prolong it, the critical interest wins out.

In general, however, we may be inclined to say that, all things being equal, critical interests are to be preferred assuming one's life is going well enough that for one to say she would be better off dead would be a foolish claim. All persons have a critical interest in seeing that their lives continue at least as well as reasonably possible. While this critical interest is being served persons have the ability to engage in activities geared toward enhancing, meeting and fulfilling their experiential interests, seeking pleasure and avoiding pain and generally seeing to it that their lives are on the whole as pleasant as possible. Were critical interest in continuing to live not being met, meeting these experiential interests would be impossible (I cannot enjoy my life if I am not alive).

What the Extended Mind can tell us about the Dresser/Dworkin Debate

Dresser and Dworkin begin defending their respective positions from similar assumptions but come to draw different conclusion. Both Dresser and Dworkin consider the mind of the patient at length as the basis for the authority of advance directives. However, while doing so they both stick to an orthodox view of the mind as a closed, first person system in which the mental states of the patient are only directly known to the patient herself. Dresser does this by

addressing the epistemological and metaphysical uncertainty that comes from being an outsider trying to divine what the patient would have, or currently does, wish, and Dworkin does so by focusing on the interests (both critical and experiential) of the patient. Both of these approaches are important parts of the story, but they are not the whole, or only, story. If, as is claimed by defenders of the extended mind thesis, the mind is something that exists, in part, outside the barrier of skin and skull, our knowledge of the wishes and interests of the patient may be far more reliable than orthodoxy suggests.

If the mind of the patient is not a privileged system only observable from the first-person point of view how we view advanced directives and proxy decision makers changes in a number of ways.

1. If external entities can be, or are, part of a person's cognitive system it is possible that a person continues to exist, in some way, after the biological substrate of her person has ceased to function properly; perhaps she even exists after her death. An advance directive is a performative act by which the patient projects her mind and wishes into the future; not in a metaphorical way, but just as literally as she if she were present in the future to make her wishes known.[27] If the contents of Otto's notebook are memories just as they would be if they were stored instead in his biological brain, an argument can be made they remain memories even after Otto's death. The effect on a reader of the words in the notebook does not necessarily change just because Otto has died. The same argument can be made for advance directives.

[27] Much more will be said of Clark and Chalmer's thought experiment involving Otto and his notebook, but it requires a brief mention here to give an argument for the above claim.

2. If it is possible for a mind to span two or more brains, or for a single brain to be separated into two different bodies,[28] then it is possible that the wishes of a single patient can be revealed through more than one human source. There is debate about whether a proxy should choose as the patient would or choose in a way that is in the best interest of the patient.[29] Though the extended mind thesis can do little to help determine what is in the best interest of the patient, it does help with answering the question as to what the patient would decide. Currently there is some necessary epistemic doubt about whether the proxy has chosen just as the patient would choose because of the complete separation of persons inherent in the orthodox view of mind.

   If, in fact, the mind of the patient has spanned two (or more) brains, this epistemic doubt begins to lift. If a patient stores her desires, memories, or what have you, in a brain apart from her own the label 'proxy' does not fully capture the nature of the relationship between the individuals in question. In such cases the person 'deciding' for the patient is not making an educated or informed guess at what the patient desires or believes is best for her, but instead revealing the decision as the patient would have. The revealing of the decision may be as authoritative as if it had come from the patient's own mouth.

   This extension of minds goes beyond just repeating what the patient once said. Disputes between loved ones over what a patient once said abound. The Terry Schiavo case is one well-known example. Schiavo's parents claimed she said would want to continue artificial nutrition and hydration, while her husband (who became her ex-

[28] The latter is discussed at length in chapter 2 as an extension of Derek Parfit's divided brain thought experiment from Reasons and Person (1984).
[29] These interests might be, or maybe often are, the same thing, but this is not necessarily the case.

husband during this ordeal) claimed she would not want her life prolonged in this manner, and it is entirely possible they are both right. At different times in her life Schiavo may have said different things to different people, as her mind may have changed or she may have made some unreflective, off the cuff comments which were taken by others to be more serious or thought out then they actually were. The question is, what would the patient say, were she able, right now? Our current understanding is that if the patient cannot communicate her wishes, and no explicit written directive exists, we are left with what an informed proxy would claim to be the case. However, two persons in the right kind of relationship might make it the case that the patient can in fact make her wishes known, through the extension of her mind into the brain of another, which in turn, uses his own body to express her choices. Advance directives on paper or statements uttered to a loved one are static and not able to account for all possible future details of the patient's situation. However, the mind having been extended into another brain means that there is a way that these new details can be taken into account and the patient can be heard after all. That extension of the wishes into another person's brain means that those wishes are still subject to thought processes and can be revised and reexamined by the brain acting as an extended apparatus. If the relationship between the persons is of the right sort, the decision made by the extendee may be the same as if they were made by the extender. (Further argument for this claim will be given in chapter 2).

3. If no directive or proxy exists what are we to do? Presumably humans have been extending their minds for as long as there have been humans (at the very least we have been doing so much longer than there has been a coherent theory of us doing so). Sometimes we perform this extension intentionally by saying to the person next to us

'hey, remember this phone number'[30] or writing a quick note so we do not forget to perform a task like calling our spouse back or picking up the book we have on hold at the library. We have long extended our minds but didn't call it 'extending our mind'. There are other times we extend our mind unintentionally (Edwards, 1999[31], Clark, 2004) in the way we organize our lives and structure the world around us. Given enough evidence and with a greater understanding of the human mind than we now have, it might be possible to derive a person's decisions or choices with a greater accuracy than we have had previously. Dementia patients who fail standard competency tests perform daily tasks better in familiar environments even though standard competency exams indicate that this should not be possible (Baum, et al, 1993, Clark, 2004). If we could 'decode' the environment in which the patient lives we might be able to learn things about the workings of her mind.

Earlier I mentioned four broad shortcomings or problems with advance directives; the issues raised by Dresser et al. which I collectively titled KPADS. These were:

1- *Change of person* (Dresser, 1986, Buford, 2008)

2- *Out of date directives* (Dresser, 1986, Bonner, 2009)

3- *Inability to cover all contingencies* (Dresser, 1986, Bonner, 2009, FitzGerald and Wenger, 1997)

4- *Change of heart* (Dresser, 1986)

There are also the four general problems which are raised by Buchanan and Brock:

[30] Much more on "remember this phone number" coming throughout this dissertation.
[31] Many thanks to Carolyn Baum in helping track down this article. It was incorrectly cited in more than one secondary source.

5- A vast majority of people have not issued an advance directive.

6- Some individuals were never competent to issue such directives (the young or those born with cognitive disabilities).

7- When an advance directive does exist, it is still necessary to find an authority who can interpret how to implement them.

8- Patients who are incompetent to a severe enough degree are no longer rightfully considered persons and thus have little to no interests to be forwarded or respected. Those interests that perhaps remain are not the kinds of interests an advance directive is aimed at protecting. The advance directive in question no longer applies because the person who drafted it no longer exists.

Given that there is a great deal of overlap between many of these issues it is expedient to combine and restate them so we can better see how the extended mind thesis lends solutions. I propose breaking these concerns into three categories and discuss each as its own chapter (or in the case of the category 'Absence' an appendix):

Personhood (1 and 8)

Does an advance directive apply in cases where the patient's condition has rendered them into a state of either permanent non-personhood (PVS or dead) or made them into a different person (severe dementia or traumatic, irreversible brain damage)? In addition to the extended mind thesis, I will turn to the work of John Locke, Thomas Reid and Derek Parfit and the study of personal identity. I assume (safely, I believe) that psychological traits are a necessary component of meaningful personal identity. This being the case, as a person's psychological traits change, so too she might become a different person all together (i.e.- metaphysically distinct person) or, in the absence of relevant psychological traits, not be a person at all.

An advance directive being part of a person's extended mind changes the temporal existence of the patient. Death, permanent unconsciousness or radical change in psychology are no longer the end of the story for a person. Personhood may outlast the brain and body since the mind is no longer dependent on these particular substrates on which to supervene.

Interpretation (2, 3, 4 and 7)

In this category we face two challenges: 1) If no proxy is named in the advance directive we have to choose the right proxy when the directive is unclear and needs interpretation, and 2) if a proxy is named, or otherwise been found, arguments may be needed as to the authority of the proxy's decision once made. It will be argued that in some limited cases the proxy is not an interpreter but a revealer of the patient's wishes. That is to say, the proclamations of the proxy are not her best guess at what the patient were to decide but rather *are* what the patient *has decided* or *would decide* and revealed through an unconventional medium. This is because of the extension of one person's mind into the brain of another.

This chapter will refer to chapter 2 where I made use of Parfit's identity arguments to make the case for the extended mind thesis. Insofar as the reader finds that argument compelling, she should also see the reasoning in the arguments in chapter 4. If the Parfit argument did not convince the reader, I still believe I can make a strong case based on what some (Clark and Chalmers, 1997, Rovane, 1998, and Nelson, 2003) have called the "unusually interdependent couple". These are couples,[32] romantic or otherwise, whose lives have so

[32] Perhaps larger groups of three or more persons are possible. Given that I believe the arguments would largely be the same, I will limit my discussion to couples. I do admit that adding additional minds into the mix could cause problems should disagreement arise; a kind of extended cognitive dissonance.

intertwined, most likely over a long period of time, that we could argue that they have come to exist as a single deliberative agent, at least for certain purposes.

Absence (5 and 6)

Here no directive exists at all. In some cases, the patient was incapable of completing a directive due to infirmity or immaturity. In other cases, a competent adult may never have gotten around to or bothered to draft a directive. Here, the extended mind thesis may give some evidence, but not as much as in cases where a directive is present. There is some overlap here between problems of absence and interpretation insofar as a surrogate is charged with revealing/determining what the patient would want or is in her best interest, but here we will assume no obvious surrogate exists. In both cases the extended mind thesis will be used as a tool for inducing the wishes of the patient from the context of the person's life and environment.

Even in the absence of a directive or a proxy we might be able to discern with some degree of accuracy a person's wishes based on the evidence she left behind in her life. External expressions of her values might be found within the way she organized and ran her life; a dusty book has long been ignored while a heavily dog-eared text with detailed marginalia was given extra attention. This idea is largely speculative on my part but not without precedent. In the field of criminal profiling criminologists are able determine some likely characteristics of an unknown suspect based on the way he committed his crimes and I will draw on lessons learned from this young, and somewhat controversial, science.

In order for this kind of interpretation of a person's desires to be a dependable tool we need a much more complete understanding of human psychology than we currently possess. However, if some version of the extended mind thesis is true we might actually leave more of our mind "lying about" than we currently realize.

Before expanding on these questions, we must first divert our attention to the extended mind thesis. The thesis has been mentioned many times but little in the way of explanation or defense has yet been given. What follows in the next chapter is an outline and defense of the extended mind thesis as one that, if not demonstrably true, is not an idea than can be dismissed outright and should be given serious attention in bioethics.

Chapter 2- The Extended Mind- For and Against

"Here is Edward Bear, coming downstairs now, bump, bump, bump, on the back of his head, behind Christopher Robin. It is, as far as he knows, the only way of coming downstairs, but sometimes he feels that there really is another way, if only he could stop bumping for a moment and think of it."

A.A. Milne
Winnie-the-Pooh

Descartes in his *Meditations* solidified in the Western philosophic and neuroscientific tradition the idea of the mind being a closed, first person entity (Ryle, 1949, Avramides, 2001, and Adams and Aizawa, 2010). Descartes is the point of demarcation in modern philosophy of when "I" became separate from my body and the creation of what Dennett (1991) calls "the Cartesian Theater." Those around me could observe the animal that "I" inhabit but only I know what was really going on inside my head, and what is going on inside my head is what makes me, 'me'. This privileged access led to a belief in a special kind of knowledge which was the foundation for all other knowledge (or so Descartes thought); cogito ergo sum – I think therefore I am. I can be certain of my own existence beyond any doubt, but the existence of any others was less certain. Our knowledge of self was beyond reproach and not in need of further defense, which is not something we can claim of any other knowledge. Nearly all theories of mind since have kept true to these ideas. Shifting this paradigm will be no easy task, but attempts have been made.

This chapter will discuss one such attempt to rethink the shape and scope of the human mind. The extended mind thesis was brought to popular light most recently by Andy Clark and David Chalmers (Clark and Chalmers, 1998) but it is not an entirely new idea (Chalmers cites

one publication from as early as the late 19th Century[33]). Though the extended mind thesis has been much discussed within the philosophy of mind, it is not terribly well known outside of it. A summary of the essay "The Extended Mind" by Andy Clark and David Chalmers follows below. This initial summary does not catalog all the nuance and minutia of the conversation, nor does it do full justice to Clark and Chalmers work. It is intended to give some grounding to those unfamiliar with the extended mind thesis. Further details will be included, as needed, as this dissertation moves forward.

Additionally, arguments against the extended mind will be considered. Notably, Frederick Adams and Kenneth Aizawa, who have raised serious concerns with the validity of the thesis. These objections concern a lack of an appropriate mark of cognition and absence of non-derived content. Though others have had important things to say about the extended mind thesis, I believe the work of these two teams of authors offers a more than adequate overview of the debate. I will discuss other commentators as needed, but for the moment I would like to add one specific author to the mix. Derek Parfit's divided brain thought experiment (Parfit, 1984), gives a nice nudge toward belief in the extended mind thesis to people who are agnostic, if not fully in doubt. As is the case with much good philosophy, Parfit walks us through a series of small steps with which we will find little to disagree about but in the end, we can draw radical new conclusions.

What I hope to contribute to the conversation about the extended mind is three-fold:

1- I will clearly define the limits and scope of the extended mind. I believe at least some of the resistance to the extended mind stems from the misconception that mind and

[33] Bradley, F.H. "In What Sense are Psychical States Extended" Mind. New Series, vol. 14 April 1895. pgs. 225-35

awareness/consciousness are the same thing, or that any and all cognition extends or is extendable. Those cognitive tasks which can extend do not threaten the "specialness" of the human mind nor do they make the human brain obsolete. I will limit the extended mind to those cognitive tasks which do not lose fidelity or accuracy as they cross over between media.

2- Perhaps the greatest sticking point in this debate has to do with the claimed absence of non-derived or original content (content which has meaning to the mind/person apart from socially established meaning) in the extended mind. Adams and Aizawa claim that the extended mind does not bear this "mark of cognition." I take Adams and Aizawa to mean that non-derived content is a necessary part of mind and that the extended apparatus is not capable of supporting or having such content. I agree, in part. In extended cognition there is still a brain to be found (cognition has to extend *from* somewhere in addition *to* somewhere), and as such, non-derived content may be found in the brain, if not in the extended apparatus.

Jerry Fodor (2009) raised as an objection to the extended mind that externalism requires internalism, but internalism does not require externalism. By this he means that if an external part of the mind were decoupled there would still exist a mind (the internal mind) of the one coupled to the external apparatus but not vice versa, ergo nothing external can be part of the mind because without the internal part there is no mark of cognition remaining in the external part. I see this as a limitation rather than an out-and-out objection to the extended mind thesis; a limitation which could make the entire proposition of extended cognition less objectionable. For the extended mind to come off there must be a part of the extended mind that is not extended, i.e.- a brain to give the external bits meaning, order and interpretation.

I will show that the 'mark of cognition' (if it even exists) sought by Adams and Aizawa does exist within the extended mind, just not in the extended parts. Non-derived content is found

within the entire extended mind qua system, but it is not necessary for non-derived content to be pervasive throughout all physical structures (either internal or external) involved in cognition.

3- I will show that there are physiological and biological grounds, not just metaphysical grounds, for accepting the extended mind thesis. I briefly examine the activities of social insects which act with a 'hive mind' and the different levels of analysis required to more fully understand these organisms. I will continue by considering how the human brain functions as a physical object which is part of a large organism. The brain, being a physical object, has extension in space, as such the activities of the mind are already distributed across time and space. Through some basic neuroscience and a variation of Derek Parfit's divided brain thought experiment I will hopefully convince the reader that the extended mind is nothing more that "stretching the wires"[34] of the human brain.

This literature review of even the broadest, most fundamental debates on the extended mind thesis is necessary. The arguments for, and nature of, the extended mind thesis are such that they are easily open to the misconception that the extended mind thesis is mere functionalism. In fact, the extended mind thesis is making stronger claims than a simple functional one. It is only with an appropriate understanding of the thesis that the reader will be able to see reasoning behind the later claims concerning the conception between the extended mind thesis and advance directives. The extended parts of an extended mind *are* part of the mind, not only functional equivalents to part(s) of the mind. It is the case that the mind extends into the world and not that the world can act *as if* it is part of the mind. Our insistence that what is extended is not part of the mind is a cranial bias without merit.

[34] This phrase is not so subtlety borrowed from Dennett (1985)

The Extended Mind Thesis- For

The main claim of the extended mind thesis is summed up nicely by Clark's phrase "the mind ain't just in the head" (Clark, 2004). The traditional physicalist view of the mind is that where the brain (or possibly the body) stops, so stops the mind.[35] The extended mind thesis, however, challenges this assertion by claiming humans can co-opt the world around them in such a way that the mind then extends past the barrier of skin and skull in ways which are variously intentional and subconscious. No one serious about the intellectual capacities of the human mind can deny that humans use the environment to *aid* in cognitive function (consider doing long division using a pencil and paper or calculator versus trying to do the calculation in one's head) but, for proponents of the extended mind thesis, this interaction is not an aid but rather just as much a part of a cognitive process as single neurons or lobe of the brain, should certain circumstances obtain. Just as neurons and brain regions serve different cognitive functions, so do external apparatuses of cognition. Those areas of the biological brain which are concerned with locomotion play little or no role in processing visual stimuli or in language comprehension. Those external apparatuses conscripted to lighten the cognitive load of memory do not function the same way as those used to solve a challenging logic problem. Not all cognitive functions can be extended. It is hard to see how you would externalize the mental tasks associated with walking or homeostatic regulation of bodily functions.

Clark and Chalmers are not making an argument about supplementing the mind but rather that some processes external to the brain or body constitute actual cognitive processes. In this

[35] I do not think the link between the brain and the mind is controversial. Of course, the 'hard problem of consciousness' (how does non-conscious matter give rise to conscious beings) is, well, a hard problem. We might be at a loss to precisely explain the link between mind and brain but for now it is enough to acknowledge and admit the link is there.

regard they are not making a merely functionalist argument. The claim in question here is not whether apparatuses external to the brain are functionally equivalent to those found in the brain, but rather that apparatuses external to the brain are part of the mind of a person just as much as neural structures within the brain itself (under the right circumstances). To write a note to remind yourself to buy milk on your way home from work is not just a heuristic to aid the cognitive task of memory. It is itself, literally, an act of remembering: an act of memory formation. The only differences between writing a note and committing the item to biological memory are location of the memory and the medium in which it is stored. If, as Clark and Chalmers say, the activity took place inside the head, it would be considered a cognitive process, then if it took place externally we should call it a cognitive process as well (Clark and Chalmers, 1998).

Is the Extended Mind Merely Functionalism?

As previously stated, there is no doubt that humans use external tools to aid cognition. We often learn to count using our fingers or by sorting small doodads into different size piles. Later in life we use calculators to balance our checkbooks. I can add 231 and 426 in my head but can do so much more quickly with pen and paper. Whether in my head, with pen and paper or with a calculator, I will arrive at the same result so long as I correctly follow the proper algorithm. If you were to ask me to add said numbers and I report the result it does not matter, functionally, how I derived that answer; you requested and received the sum. Regardless of the method employed it is *as if* I (solely) mentally did the addition.[36]

[36] Or, it is as if I used pen and paper if I added mentally. Functionality, as discussed here, is a symmetrical property. Insofar as a hammer is functional equivalent to a rock, a rock is functionally equivalent to a hammer.

At its broadest, functionalism is concerned with capabilities of a thing regardless of what it is made of. Clocks are functionally equivalent to sundials and clepsydrae, in that each is able to keep time. Functionalism within the philosophy of mind is much the same. In this specific form of functionalism mental states are identified by what they do rather than what they are made of. (Polger, 2017). Alan Turing (1950) used the imitation game to argue what it would take for a computer to be functionally equivalent to a mind. If a computer could pass the Turing Test we should say the computer is a thinking thing.

The extended mind is not as ambitious a proposal as would be tested by the Turing Test. Rather, it is about redefining, re-understanding or recategorizing what qualifies as mental; the orthodox view being that only what is internal (i.e.- neural) is capable of instantiating mental states. But can an internal mental state be functionally equivalent to an external record of that mental state? If I need to write a note to myself to remind me to pass back term papers to my students, there exists a functional relationship between this and a biological memory of the same. Clark and Chalmers would agree that this is a functional relationship, but would also take it a step farther and say this is not *merely* functionalism. The note, they would claim, is not *as if* I had mental state X but is instead a mental state in its own right.

Consider a grandfather clock. Many things are functional equivalents to such a clock: the aforementioned sundials and clepsydrae, but also digital alarm clocks, wristwatches, hour glasses and other grandfather clocks.[37] Additionally, the grandfather clock itself has parts, all of which can, in principle be replaced by a functional equivalent. A face with Roman numerals can be

[37] All these time pieces are functionally equivalent but come with different limitations. Sundials only work during the day, outside, when it is sufficiently sunny. Using a clepsydra to tell the time of day is a cumbersome task and requires a fair bit of vigilance to keep accurate time, but it is not impossible.

replaced by one with Arabic numerals. Metal hands can be swapped out for plastic ones. In neither of these cases would we say the face and hands are anything but the hands and face of the clock once replaced by a functional equivalent. Regardless of the material or typeface, these parts are performing the same functions as their predecessors and therefore have equal claim to being the same kind of thing. Swapping out parts in this manner is not (just) functionalism. The new face of the clock is just as much the clock's face as the previous one.

What if we replaced the pendulum with a sufficiently heavy loaf of stale bread? Is this aged baked good now the pendulum of the clock or acting *as if* it is the pendulum? Not only is it performing the same *type* of function, in this case it is performing the same *token* of that function.[38] The stale bread is literally the same thing as the clock's pendulum, not just a second-class, functional stand in.

We might be resistant to this radical change of parts. In the previous examples, a clock face and set of hands, the changes were material or formal (to borrow from Aristotle). But a loaf of bread as a pendulum?

This difference in sentiment results from simple chauvinism, call it component chauvinism. If it has the form of a clock face, it can be a clock face. If they are of appropriate size and shape they can be hands. To be a pendulum, it must look like a pendulum and be made of pendulum appropriate materials. A loaf of stale bread does not look like a pendulum nor can it be easily mistaken for a pendulum; therefore, it is not a pendulum. At most it is acting *as if* it is a pendulum. However, if we take a step back and compare the Roman numeral-faced, metal-handed grandfather clock with an orthodox pendulum to the Arabic numeral-faced, plastic-

[38] A sundial and a wristwatch both perform the same type of function but are different tokens of that function.

handed grandfather clock swinging a week old pumpernickel that it later became, do we see a functional difference in the whole? The change in the material and form of the parts did not affect the function of the whole. The stale bread has every bit the causal potency on the clock as a whole.

This example can be carried over to the mind. An orthodox token of a mental state is that of a state continued by, stored in/on or that which has supervened upon neural matter. If I write myself a note to remember to buy milk what difference does it make what that information is continued by, stored in/on or supervened upon. A memory written down on paper does not fit our ordinary conception of what a memory "looks like". If the written note has the same functional power as it would were it constituted biologically, it is by every measure a mental state lest we run the risk of component chauvinism. The note on its own could never pass the Turing Test, but the note coupled with the note writer almost certainly could.

Clark and Chalmers ask us to imagine Otto, a high functioning Alzheimer's patient who carries a notebook filled with important information for instances when his biological brain fails him (Clark and Chalmers, 1998). When recalling the location of the Museum of Modern Art in New York a cognitively unimpaired person searches her biological memory (assuming she has reason to believe this knowledge could be found there and that the information can be trusted). When preforming the same recall task Otto, because of his illness, needs to consult his notebook (assuming he has reason to believe this knowledge could be found there and that the information can be trusted).

Certainly, Otto is engaging in a cognitive activity insofar as he made the conscious decision to seek out the information, but why is the information stored in a biological brain a memory while what is in the notebook is not? More generally, why is the information stored in

biologically cognitive but the information in the notebook is not? Granted, the arguments at work here have a strong functionalist flavor; if process A is considered an X and process B has all the same relevant consequences of A then B should be considered an X. If an analog and a digital watch are both equally good at telling time they are functionally equivalent for that task. It is the claim of the extended mind thesis that, under *certain conditions*, such external processes and entities are properly thought of and treated as fully integrated components of a cognitive process and are themselves cognitive, not just functional equivalent. Another way to think of it is this: the analog watch and the digital watch are performing a qualitatively similar task while remaining quantitatively distinct entities. Otto's notebook is performing a qualitatively similar task but *is not* quantitatively distinct from his memory. It *is* his memory. I have two functional legs and have the ability to walk unaided. If I were to lose a leg in an accident and have it replaced with a prosthetic limb we would not be quite so inclined to say I am not walking when I ambulate but only doing something functionally equivalent to walking. I'm still walking, just using different hardware to get the job done.

These "certain circumstances" are an important feature. As with all functional relationships there are limits. A sundial is functionally equivalent to a grandfather clock so long as it is sunny, and it is pointed in the right direction. An analog watch can be used as a compass of sorts, while a digital watch cannot. Having a few lines in a well-organized notebook is one thing but standing in the middle of a library or at a computer connected to the internet does not mean the library or internet are part of my cognitive process and thus I "know" all the information contained within. What separates the extended mind from other merely functionalist relationships is the strength of the connection and character of the relationship between the

orthodox mind and the external apparatus. There are four criteria that must be met in order for an external entity to be part of one's extended mind:

The entity …

- is readily available
- is easily accessible (transparent in use)
- is a constant in one's life
- must have been endorsed in the past and there is some consequence to its future endorsement.[39]
- (Clark and Chalmers, 1998 and Levy,2007)

An example given by Clark is a wristwatch. A wristwatch is readily available when worn and does not take a great deal of physical or mental effort to use (thus is readily available and transparent in use), may be as much a part of a person's daily routine as grabbing car keys before leaving the house (a constant in one's life) and the wearer endorses it as an authority as to the time by using it to schedule her day (an endorsement with future consequences associated with it) and would be either early or late for appointments and such should the watch be incorrect. When a person wearing a watch is asked "do you have the time?" sometimes the response is to say "yes" before even looking at his watch (implying an unreflective acceptance of the watch's accuracy) (Clark, 2003). Since the effort to bring the time on the watch wearer's conscious mind takes so little effort (perhaps no more effort than recalling 5+7=12) we can say that while wearing the watch the wearer knows the time even though that knowledge is not held or

[39] This appears to be two criteria lumped together, however, as this is the way Clark and Chalmers, and Levy discuss them I will follow suit.

occurring within his biological brain.[40] By Clark and Chalmers' own admission few things meet these criteria, but some do and others may someday. Clark in particular sees great potential in the increasingly common use of smart devices such as an iPhone (Clark, 2003). As these miniaturized computers become smaller,[41] less intrusive and easier to use, we fold them into our cognitive tool kit more seamlessly. Anecdotally, consider how many phone numbers you held in your wetware 20 years ago versus today (if you carry a mobile phone). I used to know perhaps a couple dozen phone numbers by heart, while today I can recall four, two of which are no longer even in service.

Not only is it possible for humans to extend our minds to include entities in the environment, but we can extend our minds into the brains of others. This claim can be as simple as the belief that we temporarily borrow another person's brain to lessen the strain of cognitive tasks to as significant as a single person spanning two bodies and two brains.

We all have experience with utilizing others to lessen cognitive burden. If, for example, I need to remember a phone number and do not have anything to write on, I may say to the nearest person "remember these numbers" and recite the three-digit exchange to make committing the four-digit line number to my biological memory easier. Though mundane, this is a perfect example of a cognitive process (remembering a phone number) which is spanning two separate brains.[42]

---

[40] But the knowledge of how to read the watch and the desire to check the time are found in the biological brain. This is an important point related to a future section of this dissertation regarding non-derived content.

[41] Smart devices have started getting large over the past few years. There is a limit to how small a device can be and still be useable. We can only push buttons so small. Despite the increasing size, these devices are becoming less intrusive and easier to use.

[42] The fact the phone numbers are seven digits is not an accident. There is a so called 'magic number' in cognitive science: 7 ± 2 (Miller, 1956). Remembering a string of seven random

Does this phone number example meet the four requirements discussed by Clark and Chalmers, and Levy? For the first two I believe we can agree it does without too much disagreement. The brain to which the cognitive task was extended is right next to the person who originated the mental task of remembering, so it is indeed readily available and easily accessible;[43] one need only ask for it. Also, criteria four (endorsement with future consequences) is met in virtue of the fact that the task was originated in the first place. By beginning this offloading process, the originator (at least tacitly) endorsed the soundness of the extended cognitive apparatus because there was a future consequence to consider. The third of these criteria, a constant in one's life, is perhaps not as clear. Since the agent in whose brain the information extended into is not physically tied to the task originator, the two can easily be separated; she can just walk away. Perhaps this is technically true, but I do not think it is a wholly fair objection. Clark and Chalmers did not have in mind that under no circumstance could the extended part of cognition be separated out. As they say, Otto likely does not shower with his notebook, nor is it accessible while he sleeps. I think a better way to understand this criterion is to say the apparatus is 'constantly available during a relevant period of time'.[44] As it is unlikely Otto will need to access MoMA's address while showering, so too I will not need the extended cognitive apparatus involving my friend's brain to remember a phone number three

---

numbers is about the comfortable limit of most people's abilities. In addition, I suspect, the fact that we have a sort of conventional rhythm for reading phone numbers (XXX … XX … XX) is a mnemonic aid in easing this surprisingly difficult cognitive task.

[43] Though, I must admit, it is not perfectly accessible. The carrying out of this cognitive task requires the cooperation of a second agent who can, if she so chooses, be a jerk and not reveal the information when asked. I do not find this difficult to reconcile as our own memories can be 'jerks' of a sort from time to time and not be forthcoming with information we know perfectly well is found there, sitting on the tip of one's tongue.

[44] "Constantly available during a relevant period of time" is a tweaking of criteria three which will come up elsewhere in this dissertation.

weeks down the road, as at that time other avenues for offloading this act of memory will be available to me. If 'being mental" requires constant, perfect, eternal access to the content of our brains then not even our own minds are cognitive. If anything is cognitive (as critics of the extended mind thesis like to say) certainly a functioning human brain is. Why should we expect greater perfection from an extended mind?

A more pointed objection would be that this phone number example is too trivial to be enlightening and because of this triviality it fails to convince skeptics of the plausibility of the extended mind thesis. However, do not confuse that triviality of the implications with triviality of the idea. In a brief essay John Case lays out a thought experiment building on Clark and Chalmers' Otto:

> Imagine, if you will, a futuristic world containing Otto with progressively more advanced Alzheimer's disease and with his notebook replaced by successively more intelligent robots culminating in a highly intelligent robot with legal rights and a job involving its carrying Otto around to visit friends he no longer remembers but may still enjoy. Conceptually, it becomes progressively more useful to exclude Otto mostly or entirely from the boundaries of the navigation cognition required and ascribe most or all of that cognition to the robot. At least, in between, when Otto is still cognitively contributing, it would seem useful to conceptualize Otto and the robot as a team (or part of a team) with relevant distribution between them. (Case, 2004)

The cognitive task here has shifted over time from an agent to an apparatus deployed and endorsed by that agent. Otto was at some earlier point fully capable of carrying out this navigation task on his own (before the onset and progression of his Alzheimer's) and at a later point the robot was in full control. In the intervening time, however, both Otto and the robot were necessary for certain tasks while neither alone was sufficient. Therefore, the cognitive task must be distributed over the Otto/robot system in the intervening time. The same task is being performed as walking with a prosthetic limb is still walking.

Insofar as the reader agrees with Case's conclusion (that Otto was able to shift his mind, over time from his brain to his robot), I see no principled reason to require this shifting over time as a prerequisite for a mind to span two brains. Otto and the robot constitute an extended cognitive process as do two agents sharing the cognitive burden of remembering a seven-digit phone number. The only difference is that in the Otto/robot case is that there is a shifting in the amount of cognition being carried on by in any one participant at any one time with the end result being the robot doing most or all of the work at a future time, whereas in the case of the phone number, the distribution of cognitive burden is task specific and temporary. I can commit the fact the Providence is the capital of Rhode Island to memory because it will be on tomorrow's test and forget this fact later in the same day as the test. I still engaged in the cognitive act of memory. The temporary nature of the extension makes the process no less cognitive.

Derek Parfit and the Extended Mind

While studying modal logic as an undergraduate, I learned that the surest way to show something is *possible* is to show that something is *actual.*[45] I will argue here that the extended mind is possible because the mind is, in actuality, extended by its nature. The mind, insofar it is a function of, or otherwise tied to the brain, it is a spatially and temporally extended entity. If this is the case it should not be hard to see how the mind can extend outside of the skull. To show that this external extension is possible I will examine a series of thought experiments purposed by Derek Parfit. To my knowledge, Derek Parfit never weighed in on the extended mind thesis per se, but found within his work are bits and pieces which, when taken together, make a strong case for the extended mind. I will examine parts of Parfit's 1984 work Reasons

[45] Thanks to Professor Paul Graves for this.

and Persons, and form an argument in favor of the extended mind thesis. Parfit's work is an elucidating combination of contemporary medicine, speculative science and science fiction. He pushes readers through small, sometimes obvious, steps to conclusions we might not otherwise consider. I believe by examining these few disparate cases and thought experiments found throughout Reasons and Persons I can build a case for the extended mind in the same way. If we 'stretch the wires' of the mind (i.e. – expand the distances over which the brain operates) in the right way we can see how an individual mind extends and opens the door to group minds or social cognition.

Parfit asks us to consider some rather fantastic bits of science fiction given some actual facts about the way the human brain works under extraordinary circumstances. The next few sections are summaries (with some added exposition, commentary or support) of a thought experiment or medical procedure discussed by Parfit. Each of these sections will conclude with by a premise derived from the thought experiment discussed. These premises will be used to formulate the conclusion that the extended mind is possible, if not outright actual.

Where (and what) am I?

How many minds do I have? The answer depends on what exactly one means by 'mind.' The conventional view is that I have one mind, yet still feel perfectly comfortable talking about one's conscious and unconscious mind. So, does this mean I have two minds, one which I am aware and in control of and the other not? What about cases of dissociative identity (multiple personality) disorder? Can a single brain support many minds? These and many other questions like this are tackled by Parfit in Reasons and Persons. My own sense of my mind and my self is that I exist in one place at any given time. Where am I? My mind is where I am. I do not have a sense of myself being located in more than one place at a time nor sitting inside of two different

skulls nor of two persons fighting for supremacy within my own skull. I am simply "here" and I am one thing. Parfit demonstrates the mind, and therefore person, is not as unified as this first-person experience suggests. Our persons are dividable and mutable. We are not who or only what we think we are.

The Divided Brain

A medical procedure called a Corpus callosotomy[46] involves severing the dense cluster of nerves (the Corpus callosum) which connects the two hemispheres of the brain, effectively isolating them from one another.[47] Throughout both the philosophical and psychological literature (Nagel, 1971, Parfit 1984, Dennett, 1991) it is observed that the human brain can behave strangely once it has undergone a Corpus callosotomy. In a laboratory setting a demonstration may look like this:

A patient who has undergone a Corpus callosotomy is fitted with blinders to isolate each eye's field of vision.[48] The left eye is shown an image of a word, say 'jazz' while the right eye is shown a blank screen. The patient is given instructions to write his answer as well as give a verbal response to a forthcoming question. Since the image is being shown to the left eye the visual signal is being processed by the right hemisphere of the brain. As a result, the patient is

[46] This is an extreme treatment used to treat serve epilepsy to lessen the number and severity of seizures.

[47] After this procedure there still exist some tenuous connections between the hemisphere (such as the optic chiasm), meaning the brain hemispheres are not completely separated. Additional surgical procedures can sever these connections as well. While these additional 'cuts' create additional cognitive burdens for the patient, such as bitemporal hemianopia (50% loss of peripheral vision in each eye) none affect the outcome or conclusions of this thought experiment. We will assume these additional steps have been taken.

[48] For the purposes of this outlined experiment the patient is left handed. Handedness can affect the setup or outcome of these kinds of experiments, but handedness is also a factor which can easily be controlled for by switching the hand with which the patient is asked to write.

asked something like “what kind of band is it?” The verbal response given by the patient is “I have no idea” while at the same time they write the word ‘jazz’. The patient appears to have both been able to answer the question correctly while claiming ignorance of the answer. How is this possible?

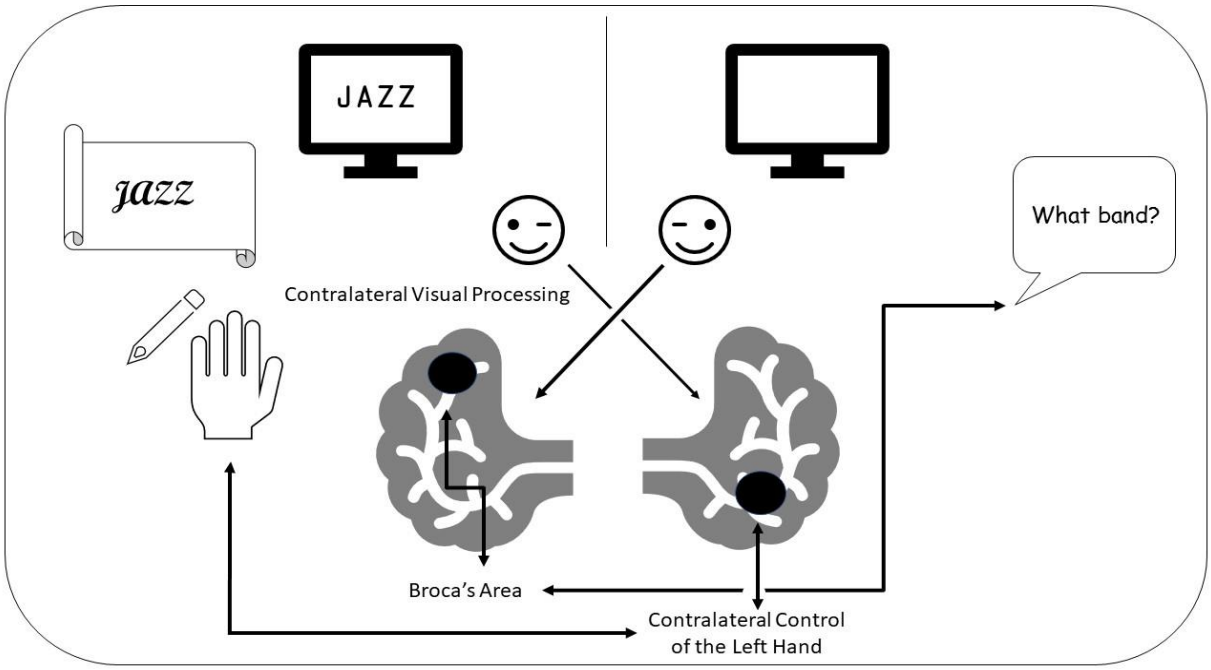


Figure 1- Diagram of experiment with post Corpus callosotomy patient

The area of the brain responsible for verbal speech, known as Broca’s area, is found, most often, in the left frontal lobe. Since the corpus callosum of the patient has been severed, the signal from the left eye of the patient has no access to Broca’s area and thus the answer is not processed into a verbal response. The area of the brain responsible for moving the left hand, however, is located in the right hemisphere where the visual signal is being sent and thus the patient is able to formulate an answer. In essence, there are two distinct conscious, cognitive

processes occurring in a single brain. Under these controlled conditions we can make it the case that the right side of the brain doesn't know what the left side of the brain is doing.[49]

This quirky phenomena of the human brain opens the door to this possibility (call it 'A'):

A- A single human brain can, under certain circumstance, be the locus of two separate consciousnesses.

I would like to urge caution here. When I claim that a brain is the center of two distinct consciousnesses, I do not wish to imply that each of us is living with a stranger rattling around in our head. There is not a creative, emotional person on the left and a rational, analytical person on the right (such distinctions are a popular, stubborn myth). Most cognitive processes are global (involve both hemispheres). When we isolate the hemispheres, each tries to compensate for the sudden loss. Since there is a great deal of redundancy and plasticity in the human brain, we might be better served referring to the "potential for multiple consciousnesses" so long as the corpus callosum remains intact. As this thought experiment plays out we will see how one brain can have more than one consciousness teased out of it.

The Missing Half

There is another rare medical procedure used to treat seizures unresponsive to other therapies called a Hemispherectomy which involves removing an entire hemisphere is the brain. A Hemispherectomy may also be performed on a patient who has suffered severe brain damage to one hemisphere due to illness or injury. If it is medically determined that the injured

[49] The controlled conditions here are important since in everyday life most people are processing sensory data with two eyes and, therefore, both sides of the brain are working simultaneously, though still more or less independently. After a period of recovery patients having undergone a Corpus callosotomy do not show many (if any at all) signs of cognitive impairment or difficulty in their day-to-day lives.

hemisphere of the brain is negatively affecting the healthy hemisphere, the damaged tissue may be removed to mitigate its deleterious effects. Persons having undergone this procedure often are able to live full lives with only some impairment (Pulsifer, et al, 2004). Most often these success stories are found when the procedure is performed on a young child since there is a great deal more neuroplasticity to be exploited to compensate for the sudden lost half of the brain, but even adults can still retain and regain a remarkable amount of cognitive function after this procedure.

What the success of this procedure shows is that both hemispheres of the human brain are not necessary for meaningful cognitive function. Memories, skills and personality can (but are not always) retained after this procedure. A person can quite literally live a life with only half a brain; which leads us to our second premise:

B- Even with only half of a brain a human being can be a cognitive agent able to engage in a wide variety of mental activities, though what activities the agent can engage in may differ from those she was able to perform while her brain was intact.

As one might expect, you cannot willy-nilly hack out half a person's brain and be surprised that there have been some changes to the patient. The effects on the personality of the patient can be severe. Even damage to a much smaller area of the brain can have radical impact on the kind of a person someone is. Phineas Gage is the classic example. Before his unfortunate encounter with a railroad spike that ripped through his frontal lobe, Gage was reported to be a happy, kind, likeable and thoughtful man. After the accident he was a mean, blowhard who was seemed unable to make plans about his future or understand the consequences of his actions. If a

Hemispherectomy is performed on a very young child (say a week-old infant, for instance) we would not expect to see serious changes to personality (since there is no benchmark to compare the child to) but when dealing with older patients the effects are likely to be more pronounced. There is no way of knowing, a priori, in what way or how much the person will change. Common psychological markers for personhood (such as memory) might be lost, but perhaps not. Gage remained the same person numerically (according to the memory theory of personal identity) but had a personality unrecognizable to those who knew him before the accident.

The Transplanted Half[50]

Although a Hemispherectomy is only preformed in rare and serious cases, there is no technical, medical reason it couldn't be performed on a healthy brain. Parfit imagines a scenario in which a person has an irrevocably damaged body (say from an auto accident) but an otherwise healthy brain. Meanwhile, two others, nearby, have irrevocably damaged brains but otherwise healthy bodies. In order to follow the utilitarian imperative that, all things being equal, we save as many lives as possible, we divide the healthy brain and place each half in one of the two healthy bodies.[51,52] Parfit believes when the patients wake up we are faced with a metaphysical

[50] Here, and elsewhere, the procedure described is often times referred to as a 'brains transplant.' My wife, Lauren, pointed out to me that we should more properly refer to the procedure as a 'body transplant' since the patient, after the surgery, is identified as the person whose brain, not body, is in question. Point taken. I will nonetheless still refer to a 'brain transplant' to remain consistent with previous discussions of the issue.

[51] By Parfit's own admission in 1984 when Reasons and Persons was written, (and 30+ years out today) this medical miracle is beyond our abilities but is nevertheless possible in principle.

[52] Since the early 20th Century attempts have been made to transplant the whole head of an animal onto the body of a conspecific (Roach, 2003 and Young, 2002) which, arguably, is functionally equivalent to Parfit's purposed brain transplant. These experiments have met with limited success. Recently a Russian man volunteered to undergo a human head transplant (the Russian volunteer is providing the head). The procedure was slated to take place sometime in 2017. (I am unsure if the procedure took place as scheduled) (Welch, 2016 and Kean, 2016). Putting aside the ethical minefield this purposed procedure creates it is nonetheless informative about the extent to which cognition is an embodied process. Two Italian bioethicists, Anto

conundrum concerning their identity. Which, if either, of these two persons is the person to whom the whole brain belonged?

Parfit claims there are four possibilities:

1- Neither is the same person

2- One is the same person

3- The other is

4- Both are the same person

Each of these possibilities has drawbacks and none, in Parfit's mind, is fully correct despite some coherent arguments in favor of each. What's interesting, as Parfit points out, is that even if we were able to perform this procedure we might still be no closer to answering this question. Both transplant recipients could claim or deny being the person to whom the whole brain belonged but we may still lack the tools for evaluating these claims.

I believe there is a fifth possibility which Parfit did not consider. This possibility is the presence of extended cognition. There are two persons, neither of whom is the same person as the one whose brain it was, but in addition there exists possible cognitive processes spanning the

---

Cartolovni and Antonio Spagnolo, (2015) point out that it is not just a matter of moving the brain from one body to another and everything cognitive comes with it, saying:

"Despite his [Dr. Sergio Canavero's (the surgeon who announced plans for the procedure)] vision, modern cognitive science shows that our cognition is an embodied cognition, in which the body is a real part in the formation of human self," they write. "Therefore, the person will encounter huge difficulties to incorporate the new body in its already existing body schema and body image that would have strong implications on human identity."

I find myself both repulsed by the ethical issues at hand here and intrigued by the medical and metaphysical implications of this procedure should it take place and be successful.

two persons which are the cognitive processes of the brain donor. There are some cognitive actions which will require both persons to complete (since, let's say, person A has retained the neural tools for reconstructing experiential memory X and person B has retained skill Y). We have two persons, yet three instances of cognition.

Here we can draw a third premise:

C- In theory, a single brain can be separated into two parts, transplanted into healthy bodies and retain meaningful, conscious, cognitive function.

As admitted a brain transplant is not yet medically possible, but not theoretically impossible either. Premise C gets it strength from the fact that if a person has a healthy, whole brain, that person can (possibly) live and function with half of that brain removed (and a lengthy recovery with much physical and cognitive therapy). It does not matter which hemisphere is removed. The outcomes will be qualitatively different but in the end whether it is the right or left hemisphere remaining we may still have a conscious, cognitive person.

<u>The Argument</u>

Here are our premises so far:

A- a single human brain can, under certain circumstances, be the locus of two separate consciousnesses.

B- Even with only half of a brain a human being can be a cognitive agent able to engage in a wide variety of mental activities, though what activities the agent can engage in may differ from those she was able to perform while her brain was intact.

C- In theory a single brain can be separated into two parts, transplanted into healthy bodies and retain meaningful, conscious, cognitive function.

The transplant case causes a conflict with our intuitions. On the one hand, we don't want to say that the donor person no longer exists. After all, the brain is still functioning, albeit in pieces, and is interacting with the world via its new host bodies. As far as we believe that the brain is 'where' a person 'is' or, at the very least is, intimately tied in some way to personhood or personal identity, we will want to insist that the donor be somewhere rather than not exist at all if the psychological markers of personal identity survive the process since the brain (which still exists) is the substrate for these psychological makers. But we also have an intuition that a person, and therefore her cognitive processes, must be located in *a* place, not *multiple* places. "I" cannot be "in" *this* body and *that* body at the same time.

However, this misses a simple fact of human anatomy: the brain is not a geometric point. The brain is a physical object and, as such, has extension in space. Therefore, cognition qua physical process takes place over time and across distances. Consider the act of ducking from an object hurtling toward your head at high speed. From your point of view, you appear to see the object and *then* duck almost instantaneously. Why would it be otherwise; doesn't cause precede effect? You certainly wouldn't duck if you didn't think there was a reason to. Whether the decision to duck was a reflex or a conscious split-second choice is not relevant here. What matters is that from the point of view of the one who ducks the object was sighted and action was taken to avoid being hit, in that order. In fact, strange as it sounds, this is the opposite order of how things actually happen.

The areas of the brain responsible for reaction to visual stimuli are located in the front of the brain, closest to the eyes. As a protective mechanism this makes a great deal of sense. The

sensory information needs to travel the shortest possible distance, thus saving precious milliseconds which could be the difference between keeping and losing an eye, or even life and death. The areas of the brain responsible for conscious awareness of visual stimuli and conscious action are located much farther back in the brain. The gap perhaps spans 6 inches or so but accounts for about an 80-millisecond delay (due to the distances between neurons and the speed of the electrochemical processes) in exactly when each brain region initiates its processes. Though it is technically true that the object flying was "seen" and reacted to, it was not seen by you insofar as "you" are your consciousness. The unconscious process of detecting a threat and reacting to it was started and stopped before you were ever consciously aware there was a problem. An 80-millisecond delay is not long enough for us to notice so the story we tell when recalling events places our conscious awareness of the threat before our defensive action. This perhaps fits with our conceit that we are in control of ourselves or squares with our day-to-day understanding of cause and effect, but in fact this story is wrong. The cognitive processes related to the recognition, response and understanding of a threat take place at different times and place. The temporal and spatial locations of these cognitive processes are in close proximity, yet gaps of time and space do exist.

This being the case, I would like to make the argument that once the brain has been divided and transplanted (as per our above thought experiment) unified cognitive processes (at least some of them) involving both halves of the brain are possible; the only differences being the time needed for these processes to occur have lengthened due to the increased distances and the means of transmission are no longer neuronal connections but rather external media (verbal communication, visual cues, etc.) which also contributes to the longer timeframe.

Let's consider a variation of our post Corpus callosotomy lab experiment. Instead of isolating the fields of vision in a single participant we will substitute in our brain transplant recipients. We'll call the patient with the left hemisphere 'Lefty' and we'll call the other 'Wayne'. We will assume that the transplant has left each with some non-serious cognitive impairment, but are each functioning human persons.

Each recipient is placed in a room with a monitor. Lefty is in the room on the left and Wayne is in the room on the right. Lefty and Wayne are not able to hear or see each other. Both will be shown some images and asked some questions. Wayne is shown the word 'dollar' and Lefty is shown a blank screen. The question is asked of both "what kind of bill is it"? Lefty naturally says there's nothing on the screen while Wayne replies "a dollar bill."

So far everything is as we would expect. Lefty and Wayne are separate persons (we might not be exactly sure what person each is, but they are each some person or other) and as such, gives answers based on their current, isolated experience.

Now we'll try again with a few tweaks. Rather than being totally cut off from one another Wayne can speak to Lefty and Lefty can hear Wayne, but not the other way around. They are told they have to work together to solve a simple arithmetic problem, but each will only have part of the equation. Wayne is shown the multi-variable equation "X + 5 - Y =" and Lefty is shown the numbers 7 and 6. Wayne knows the operations and one addend. Lefty only knows that there are at least two numbers involved in the equation but has no idea what to do with them. All Lefty can think to do is say aloud "I've got a 7 and a 6 over here." Since Wayne knows the basic rules of arithmetic he realizes it does not matter which number is filled in for which variable and soon calls out the answer, "6".

Again, nothing too surprising here. Any two people in possession of whole brains could have done exactly the same thing. So let's make some more changes. They can speak to and hear each other. They need to come up with the answer through any means available. Lefty has pen and paper, Wayne does not.

Both Lefty and Wayne are showed the equation "7 + (5 – 2) + 3 x 3 =". Both Lefty and Wayne realize the order of operations here matters. They remember learning it in school. However, Lefty doesn't recall the rules for order of operations, but Wayne does. The problem is that since losing half of his brain Wayne's memory isn't what it used to be. He can get as far as the first two operations but can't remember the sub-conclusions long enough to apply the third. The dialog begins:

Lefty: "I know the order here matters but for the life of me I can't remember where to start."

Wayne: "That's OK. I know where to start but I can't keep track of things long enough to get the final answer."

Lefty: "Tell me what to do first, I've got a pen and paper in here."

Wayne: "Alright, first do the equation in the parentheses."

Lefty: "OK, the answer is 3"

Wayne: "Now do the multiplication."

Lefty: "That gives me 9"

Wayne: "Now just work from left to right."

Lefty: "The answer is 19"

Unlike the previous two scenarios where we could have substituted in any two people possessed of whole brains and achieved exactly the same results, the third case is importantly

different when we consider the origin of the persons involved. There was a single person in the past who was possessed of all the knowledge necessary for completing this task and it is that person's *tokens* of knowledge that are at work here, not just qualitatively identical knowledge of the same *type*. This is a single brain preforming a single task over a distance greater than the Corpus callosum is wide.

We can consider a post Corpus callosotomy brain in a couple of different ways. We can say it is one brain separated between bodies or we can say it is now two distinct brains. Either description make sense in the right context. But here the context makes it a single brain. The knowledge and skills needed for the task are dispersed across the hemispheres as would likely be the case in a unified brain and they were once a unified brain. To solve this single arithmetic problem each hemisphere needed to draw on different past experiences from the brain had when it was a unified whole. Only by acting as a unified whole again was it (the brain) able to solve the problem. Two whole brains having had different histories, and pulling for disparately acquired knowledge bases, are not using the same cognitive resources but instead are pooling each's resources together.

Aside from the difference in histories, an important difference between a brain as a unified organ in a single body versus a split organ spanning two bodies is a concern over the immediacy and directedness of the connection. There are concerns as well regarding the means by which information moves between a whole brain and a split brain (actions across neurons versus external media such as verbal communication) but these are more differences in degree, not of kind. If in an intact brain the right hemisphere needs some information the left hemisphere has (or more accurately, is able to reconstruct), signals are sent across neural connections. This electrochemical process, as argued, takes place over time and space but it is

still a process contained within a skull and occurring within a single organ. The split-brain persons do not have this same connectedness. The mode transmission of the information is not electrochemical, it is through sound waves if speaking and visual processing if written. However, an intact brain can engage in this non-neural information conveyance as well.

When a person with a whole brain speaks aloud while thinking through a problem the act of speaking is doing real cognitive work. When one speaks, one hears. By speaking out loud the thinker is engaging the auditory areas of the brain so as to bring that processing power into play (Dennett, 1991, Lupyan and Swingley, 2011). The same goes for working problems out on paper or a blackboard in getting the visual centers of the brain up and running. In the case of the divided brain the speaking/hearing or doodling/seeing are performing a similar task as they would in an intact brain. With the brain split between two bodies there is more necessity to the externalization of thoughts (meaning one does not *have* to talk out loud to solve a problem alone but when working with others some kind of externalization will be required since the other brain is not neuronally connected) but this externalization is nonetheless activating the brain areas as it would if the brain were intact. Lefty needs information that Wayne has. A verbal request (as opposed to a neuronal signal) is sent, Wayne's brain is activated and a verbal utterance containing the relevant information is sent in response. With our intact brains we extend or cognitive processes through the air to our own ears. Lefty and Wayne are doing the same kind of thing, though perhaps losing some efficiency compared to an intact brain.

Is this Really an Extended Mind?

Though Lefty and Wayne have solved a complex problem using cognitive resources separated by a distance greater than the size of a skull, it might still be argued that this thought experiment does not describe the extended mind as defined by Andy Clark and David Chalmers,

and others (Clark and Chalmers, 1997, Levy, 2007 and Case, 2004). If we are just 'stretching the wires' as it were, would it be a case of extended cognition if we merely enlarged the skull and pushed the hemisphere apart with a modified Weitlaner. It is, after all, still a single brain we are dealing with. The claim of a cognitive process existing between the two half brains gets it strength from the fact that the halves were, in the past, a unified whole. Memory X having been reconstructed in one half of the brain is not a memory of the same type or content as a memory reconstructed in the unified whole brain in the past. It is literally the same token memory,[53] but now accessible to the separated half though externalized requests for information. If I have the memory that Abraham Lincoln was born on February 12, 1809 and you have a memory that Abraham Lincoln was born on February 12, 1809 our memories have the same content but are totally separate tokens of the same kind memory as this memory results from reconstruction process taking place in numerically different brains. In the split-brain case this is not so. The memory of Honest Abe's b-day is the same token memory before and after the severing of neural ties; which is different than extended cognition in which one brain uses a separate brain (one it has not been biologically unified with) to store information for later use.

Consider again the phone number example. You may say to the nearest person "remember these numbers" and recite the three-digit exchange. You have coopted neural matter other than your own into your cognitive process. The brain of your assistant is doing cognitive work (remembering) not just acting as an aid to your cognition. Though mundane, this is a perfect example of a cognitive process (remembering a phone number) which is spanning two separate brains. You had in your wetware a memory of the 3-digit exchange (at least long

[53] Here, and forward, 'token memory' is meant to refer to the token ability to reconstruct a particular memory.

enough to relay it to someone else) and now the person next to you has a memory of the same content. This is not what has occurred in the divided and transplanted brain case. The memory in the transplant case is not a copy, it is the same memory as it always was.

In order to resolve this conflict, we need to consider a couple of things: 1- when exactly did cognitive extension take place when the brain was divided and transplanted? and 2- what is the origin or purpose of coopting another's grey matter? Once these matters are attended to, I will argue that the phone number case and the divided brain are similar enough in the psychological continuity that exists in each that they both should be counted equally as instances of extended cognition.

Setting aside the above argument that minds are extended due to the nature of intact brains, the soonest we *could* principally say extension occurred was at the moment of hemisphere separation, but I do not think this is the case. Most memories are dispositional at any given time and are not brought into consciousness until either they are intentionally recalled or involuntarily brought to mind because of one's current circumstances (e.g. – a pleasant smell wafts one's way, stirring up fond memories of Grandpa's Angel Wings and rosettes).

When Lefty and Wayne found themselves in separate rooms they did not know exactly what was going to be asked of them. As such, Wayne wasn't likely thinking about the order of arithmetic operations, nor Lefty the derived meaning of symbols such as '+', 'x' and '7'. If it were not for their contrived circumstances, Lefty and Wayne may have gone the rest of their lives never having such thoughts come to conscious mind and the same can be said about the brain were it still undivided.

Before the math problem was given to Lefty and Wayne we did not know the totality of what memories or skills were to be found in each separated hemisphere, but we can say that the

once unified cognitive machinery (i.e. – the brain) has been extended some memories and skills will be found in one place and others in others. Not until the attempt to solve the multistep math problem has begun do we see an extended cognitive process. The machinery being extended does not mean the cognitive process is extended (no cognitive process exists until it has been initiated), only that *if* certain cognitive processes are initiated they will be extended cognitive processes.

Once Lefty engages with Wayne (who is now a numerically different person than Lefty) the extended cognitive process begins. At some point in the past many of the psychological characteristics which now define Lefty were part of a person who has knowledge of the order of arithmetic operations. There existed psychological continuity between this knowledge and those psychological markers found in both Lefty and those found in the then unified brain. In some sense a mathematic memory of this type still belongs to Lefty though the token memory does not.

When I ask you to remember part of a phone number and you successfully do so a token memory of the three-digit number (let's say it's '853') is created. But this is not the only way token memories can have the same content. Say there is another person, Paul, nearby when I make my phone number request. Paul is ruminating on how many coins are in his (apparently ample) pocket, 853 or remembering his gym locker combination '8-5-3'. The contents of Paul's memory and your new memory are the same in terms of content, but they are semantically different, each with a different origin and purpose for which the memory was formed. Though you and Paul both have memories with identical content you can make no claim to Paul's memory being part of your extended cognition. The origin of the content matters. There is no psychological continuity between myself and Paul, whereas there is such continuity between you

and me. I held the number in my memory (at least as long as was needed to give it to you), made the request of you, relayed the information and you formed the new memory. I played a causal, intentional role in your memory formation as a means of engaging your brain into my cognitive process.

Lefty and Wayne both played a causal role in the formation of the memory of the order of operations (insofar as both hemispheres were somehow involved in the memory being formed), thus psychological continuity exists here as well. These cases of psychological continuity make the extended mind claim plausible. No such continuity exists with Paul; the matching of the contents of the memories is a coincidence.

Lefty and Wayne are just as much separate persons as you and me, but the psychological continuity in both pairs is what makes the case for extended cognition, not the historical relationship between the brains in question.

Conclusion

Parfit's experiments with the question "where a person is located?" lead him to some strange and sometimes counterintuitive answers. By the above reconsidering his thought experiments we have opened the door to, and generated evidence for, another possibility: the extended mind. Insofar as the mind is contained in, made up of, emerges from or supervenes on physical processes of the brain we can say that, in a real sense, cognition is already extended; it's extended in the head. Essentially, the above argument in favor of the extended mind doesn't start from the claim "the mind ain't just in the head" but instead "the mind ain't just at a single point'. The mind has always been extended. Clark and Chalmers, et al. are just stretching the wires.

Additionally, I have attempted to show the parallels between Parfit's divided and transplanted brain and more conventional conceptions of the extended mind. What is more important than the history of the brains involved is the psychological continuity that exist within a brain and between brains.

The Extended Mind Thesis- Against

If Clark and Chalmers's "The Extended Mind" is canon for those in support of the extended mind thesis, then "The Bounds of Cognition" (both the article and later, book) by Frederick Adams and Kenneth Aizawa is the same for those opposed. Adams and Aizawa strongly disagree with the extended mind thesis from a realist position, yet admit "these mechanisms [meaning those mechanisms involved with cognition] could (conceptually, metaphysically, and physically) occur outside the brain, they typically[54] do not" (Adams and Aizawa, 2010). Adams and Aizawa argue the widening spread of belief in the extended mind thesis can be attributed to what are, in some cases, some pretty basic fallacies in argument.

Adams and Aizawa dismiss the extended mind thesis in two ways, one having to do with the apparent lack of a "mark of the cognitive" and the second that they call the coupling/constitution fallacy. In discussing the mark of cognition, Adams and Aizawa emphasize the importance between distinguishing those things and processes which are cognitive from those which are not. Without a proper definition of what constitutes a cognitive process, they say, it makes little sense to talk about, or defended, a theory of extended cognition. They

[54] I have had trouble squaring this claim with the rest of what Adams and Aizawa have said. They are denying the possibility of the extended mind but here say it is not 'typical.' Clark and Chalmers are themselves careful to point out that instances of the extended mind might be rare, so at least on this point the two sides agree. Since the rest of Adams and Aizawa's work is focused on denying the extended mind entirely I will not question in what cases they might think it possible.

believe what separates the cognitive from the non-cognitive is the presence of original or non-derived content. As for the coupling/constitution fallacy, they claim supporters of the extended mind fail to heed some basic distinctions such as that between a cognitive system and a cognitive process and between a causal and constitutive relationship. Each of these objections will be handled in turn.

I do not believe either presents an insurmountable problem for the extended mind thesis, though it will require some tweaks to our understanding of the theory. We will need to reconsider the role and (literal) place of original or non-derived content in cognition. Adams and Aizawa fail to acknowledge the fact that not all instances of or steps in a cognitive process require or contain non-derived content. Further, I will question the distinction between cognitive systems and cognitive processes which Adams and Aizawa allege. I do not believe such a distinction exists. I will argue that a cognitive system is anything that supports a cognitive process and cognitive processes are the parts that constitute a cognitive system. The system and the process are the same thing and one cannot exist without the other.

The Mark of Cognition

Adams and Aizawa rightfully claim: "It is not helpful at all to be told that cognition extends without some idea of what cognition is" (Adams and Aizawa, 2010). Cognitive processes by virtue of being cognitive processes (must) have certain characteristics. It may be difficult to tease out what these characteristics are, but Adams and Aizawa are certain these characteristics are lacking from those entities involved with extended cognition. The need for a mark of cognition arises from the need to separate entities into two groups, those with are cognitive processes and those without or are not. The typical descriptors of cognitive processes are often too broad to be useful. If we define a cognitive process as information processing, we

would have to conclude that CD players, thermostats and pocket calculators engage in cognitive processes.[55] If we say merely any dynamic process, "then … cognitive processing will also be found in the swinging of a pendulum of a grandfather clock or the oscillations of the atoms of a hydrogen molecule" (Adams and Aizawa, 2008) and thus we run a similar risk. Information processing and dynamic process are perhaps necessary but Adams and Aizawa do not seem to believe that they (together or separately) are sufficient for defining what a cognitive process is either. By considering the presence of non-derived content as a mark of cognition we avoid these problems of broadness. There is a very small subset of processes which can plausibly be said to have non-derived content. Original or non-derived content is mental content that does not derive its meaning form an outside source. This kind of content is intentional, meaning it is content *about* something. It is content that is meaningful to the person, not just syntax and does not derive this meaning from outside or social forces. I have found few useful examples of non-derived content provided by other philosophers. It is discussed as if it is a sentiment or disposition toward mental content rather than a set of token contents. So as examples I will throw in:

- When I am hungry the drive to eat is not derived. The feeling of hunger has meaning to me.
- If I am playing a game the desire to win is not just a result of following the rules (the syntax of the game). I want to win because I want to win; because winning means something to me.

[55] Which, to be fair, is a bullet some are willing to bite. There are plausible readings of Chalmers (1996) which leads to this conclusion.

- The love I feel for my son is not captured by the dictionary definition of the word 'love'. It is just the way my son makes me feel.

These mental contents, and countless more, are just part and parcel of being a human being. We do not need to be explicitly taught what everything means. Some meaning is just part of who we are.

Consider, by contrast, the symbol '@' which has a well-known (derived) meaning in today's email heavy society. We see the symbol and read it as saying 'at' in the same way we would if we saw the characters 'a' and 't' next to each other. This meaning of this symbol has only come into use recently. I remember sometime after graduating from high school in 1993 asking my friend Bill what this strange character on my keyboard meant. The earliest usage of '@' dates back to the 15th Century and since that time the symbol has meant 'amen', been an abbreviation for a unit of weight (Italian- 'arroba') and denoting a particular volume of wine in addition to its modern meaning and numerous uses in computer programing. This *is* derived content. There is nothing about the symbol itself that necessitates it means one thing or another. Its meaning it purely conventional; just as, depending on your societal view the swastika is either a symbol of hate and the Nazi party (post World War II) or meaning 'auspiciousness' (in various Indian religions). Written (and to a *somewhat* lesser degree spoken[56]) language is also derived content. The same combination of letter can mean drastically different things between languages using the same alphabet. In English 'after' is a preposition meaning an event occurring subsequent to another, while in German 'After' is a noun meaning 'anus'.

[56] Some vocalizations such as non-linguistic screams of pain (as opposed to 'ouch', which is a learned word) or a crying baby appear to have an innate meaning or, at least, we have an instinctual reaction to them.

The requirement for non-derived content has other supporters within the philosophy of mind, notably John Searle and his 'Chinese room" thought experiment[57]. There is lengthy literature regarding the importance of non-derived content as it relates to the Chinese room (Block, 1978, Searle, 1980, Block, 1981, Searle, 1984, Churchland and Churchland, 1990) and the debate continues. The point Searle is trying to make is that we cannot get to intentional meaning through analysis of syntax alone. No examination of Chinese (or any other language, written or spoken) will be enough for one to "know" Chinese if the analysis is divorced from how that language functions in the world. Looking at the string of characters 'A, P, P, L, and E' cannot come to mean the fruit of a particular set of species of tree if all I have to learn from is a set of characters and a rulebook.

By this reasoning Otto's notebook lacks meaningful (i.e. - non-derived) content, since words on paper have no meaning apart from the meaning we give them, and is thus the notebook is not cognitive. Otto was only able to make notes in his book because of the derived meaning of written language. Even if he were using some private language which only he knew, we would still have to call the words on the paper derived content since Otto needs to give meaning to the symbols (since symbols, by definition, have only derived meaning). Otto certainly engaged in a cognitive process when choosing to write the location of MoMA in his notebook. He also engaged in cognitive processes when reading the notebook and acting upon the information read. But the fact remains that the words on the paper do not, themselves, have non-derived content, and thus are not cognitive.

[57] The broad picture of the Chinese room is well known throughout philosophy (I have never brought up the idea with a philosopher or computer scientist who was not aware of it). Rather than rehearsing the details of what has already been said by those philosophers cited (and many others) I will offer a novel argument concerning non-derived content which is specifically geared toward its role in extended cognition.

At this point, there are three approaches that can be taken in objection to Adams and Aizawa's insistence on non-derived content as a mark of cognition. The first is to demonstrate there is no such thing as original content.[58] If there is no such content, then it makes no sense to claim it is a mark of cognition. Secondly, we could argue that non-derived content exists but is not necessary for, nor plays a role in, cognition (i.e.- non-derived content is epiphenomenal) which seems as unlikely to me as it is unappealing. Or lastly, we can attempt to show that the extended mind does indeed have non-derived content and thus carries the mark Adams and Aizawa seek.

I will assume for the sake of argument that non-derived content does exist, but add that non-derived content is not necessary for all forms of cognition; a position with which Adams and Aizawa agree. Consider a common task such as driving a car. Certainly, there is attention and meaning going on (watching for pedestrians, reading street signs, a desire to get where one is going etc.) but are we really consciously aware of everything going on around us? In reality we are only aware of a small portion of the data our brains are processing at any moment (Dennett, 1991, Gangopadhyay, 2010). The physical make-up of our brains creates limits to our conscious awareness of our surroundings and, yet we are able to seemingly respond as if we are perfectly conscious of what is going on around us. When an object is thrown at your head, you dodge. You may believe you did so because you made a decision to move rather than be hit (the details of the phenomena are discussed in the previous section in conjunction with the Parfit divided brain thought experiment).

[58] Two very good arguments for this position can be found in Dawkins (1976) and Dennett (1990).

Our brains and bodies are capable of performing important tasks without us, so to speak. When we speak we do not have to consciously search for each and every word. If we are uncomfortable in our chair we can shift position without even noticing. When processing auditory language we "hear" distinct words despite actually hearing anunbrokenchainofsounds. Our intentions are not always the drivers of what seems like our conscious actions. However, since I am granting the existence of original content, it will not be necessary to argue cognition is possible without it. This being the case, I will explore the third of the above options to make my case in favor of extended cognition. Non-derived content plays a role in cognition insofar as a brain is a necessary component for cognition and brains contain non-derived content.[59] As with any other process or system, cognition is made of parts or stages. I am limiting extended elements of cognition to those parts which are concerned with objective information or processes only (that is to say information and processes which loses no fidelity as it moves across media). I do not think there is any reason a priori to say that cognition can never extend in such a way as to encompass or contain original content (radically new technologies may make this the case) but for the time being, insofar as original content plays a role in extended cognition, that content is found in the brain of the person(s) cognizing and not in the extended apparatus. Otto's notebook contains much information. The sentence "the Museum of Modern Art is located at 11 W. 53 St." only has meaning so long as someone wrote it to convey that meaning and there is someone to read it, be it the author or someone else. Adams and Aizawa argue that cognition is an intentional activity and thus necessitates original content. Not all steps in a cognitive process,

[59] I will remain agnostic on the question of whether or not some sort of manufactured intelligence (MI) may someday be able to instantiate non-derived content. I have no reason to believe at any such MI exists at present.

however, involve original content or intentionality, but rather only certain steps which take place in the brain.

Where Non-Derived Content is Found in the Extended Mind

Adams and Aizawa claim the extended parts of the extended mind lack a mark of cognition. This begs the question: do all aspects of cognition bear this mark? If the answer to the above question is 'no' we can further ask: if not all aspects of the orthodox mind bear this mark, is it possible for extended elements to be part of cognition even if these elements lack this mark? I will continue to use the example of Otto's notebook in discussing necessity (or lack thereof) of original content in mental processes. Since Clark and Chalmers' Otto thought experiment deals with the possibility of memory being extended I will, for the time being, limit my discussion to that mental phenomenon. Here, memory and recall stand in for any mental process. Not all things which occur in the mind are memories (desires, emotional states, hallucinations etc.) but the reasoning which follows will hold true for other mental phenomena as well to the degree those phenomena are extended.

Adams and Aizawa claim the contents of Otto's notebook do not qualify as mental phenomena or content because it lacks non-derived content. This argument is persuasive. Words on a page are open to interpretation and shades of meaning depending on who the reader is. If the reader has no knowledge of the language in question the words written in there will *not* have any meaning at all *relative to that reader.* Memories stored biologically are subject to the interpretation of their 'owner' but in many (but not all) cases have a semantic element by virtue of being made in part of non-derived content. But, if we look at how the words got into the notebook we see things differently.

Consider the way in which events outside the mind come to be memories and later recalled. We will consider the example of Otto coming to have the memory of the address of the Museum of Modern Art and recalling it at a later time. Before the onset of his progressively worsening dementia Otto is not yet cognitively impaired and is thus relying solely on his biological wetware for acquisition, formation and retrieval of memories. There are a number of steps in the process of committing a fact or event to memory. Endel Tulving (1984) describes the process thusly:

1- *Original Event*- Simply, this is the event or episode that is remembered. In this case the event is the set of circumstances which caused Otto to form the memory of the address of MoMA.

2- *Encoding*- "…is the process that converts an event into an engram" (Tulving, 1984). This is the biological process (whatever it turns out to be) by which an original event comes to be a memory. Whereas, before, Otto did not know the address of the museum, he now does.

3- *Engram*- This is not a process in memory formation but rather the result of encoding. This is what is searched for or recalled when retrieving a memory of an event. An important feature of engrams is that they are mutable as demonstrated by the fact that our memories change over time. Until engrams are retrieved or accessed by some retrieval process (conscious or unconscious) they are mentally inert [60] (Tulving, 1984). The engram in Otto's brain encoded for the address of the museum is sitting there, doing nothing, until some retrieval process accesses the engram.

[60] Here 'mentally inert' means semantically inert. An engram is purely syntactic until it is accessed and put into working memory. It is just information in the form of a brain state just as braille is information in the form of raised bumps on paper.

4- *Retrieval*- This is the process of bringing an engram into a working memory space. This process could be conscious, as in actively trying to remember the address or unconscious as when a familiar song is heard reminding one of a fond event.

5- *Ecphory*- This combining of the retrieval cue (above the conscious attempt to recall the address or the fondly remembered song) and the engram into a conscious or unconscious mental state.

Though we might differ on the details about how this process unfolds, the difference will not affect the overall picture. The above, or something very much like it, is how memories are formed and recalled. I think this is perhaps the shortest, adequate model of the memory process so for its simplicity and clarity we will stick to Tulving's model. I do not think a model with more detail will be any more or less useful for our purposes here.

What is important to notice here is that all of the processes (excluding the outside event which is the subject of the memory) took place inside the brain. This process of remembering is an unconscious process of encoding and storage (I do not choose what neurons or chemical compounds come into play). We can, of course, consciously commit something to memory by repeating it or concentrating on it, but still the nuts and bolts of the memory formation inside the skull remains an unconscious process.

Contrast this scenario with the following:

We have a sophisticated machine that is able to functionally duplicate Tulving's memory process in a contrived and controlled way. I have in mind here a sort of robot with cameras taking in visual input. Let's say that the robot's task is to look at a series of symbols on a wall (we'll use %, $, and #). The robot "sees" the symbols and depending on their order will execute a

particular command when the lead programmer enters the room.[61] If the symbols are ordered % -$- # the robot will write this order into memory and upon seeing the researcher the robot will use a voice synthesizer to utter MoMA's address. If the symbols are ordered # -%- $ the robot will give the address of the White House. All other combinations will cause the robot to sit next to the door and await further instructions.

Although we can shoehorn the activities of this robot into the steps of memory formation and retrieval as described by Tulving, we would be reluctant to say this is a case of memory in the sense in which we apply the term to humans. The reason for this may be in Adams and Aizawa's claim that the mental must bear a mark of cognition, non-derived content, which the above robot scenario does not.

The important distinction has to do with the mind of the agent doing the remembering at the time of the original event. For Otto, this was about something, that is to say, intentional. For the robot it wasn't about anything. But the extended mind is concerned with what happens when the information moves from biological memory to an external apparatus. Once this happens has the "memory" retained its original content or has it become just another bit of derived information more closely related to the allocated hard drive space of the robot?

To say that the contents of the notebook lack original content is a temporal claim, the truth of which depends on the exact circumstance under which the contents came to be written down and the circumstances under which it is observed. In the above case of the address reciting robot we see something analogous to a person's mindless doodle in a notebook. There is no meaningful content; the "memory" was made with no more thought that the squiggle was

[61] We'll assume all additional problems of facial recognition by computers have previously been solved.

squiggled. There was never any non-derived content because the inscribing was never *about* anything in the first place.

However, one thing that is key here is that the information now found in the notebook was once found in Otto's biological memory and thus, there was a time when the information contained in the memory did bear the mark of cognition. This would be true even if the information to be remembered was just "passing through" Otto's brain in transit to the notebook. If Otto had no intention of committing the information to his wetware because he was writing it down (like jotting down an unfamiliar phone number for later use), it still had to pass through his brain and become a written task which was *about* something. Once the memory was transferred into the notebook Adams and Aizawa want to say that the mark of cognition is lost. However, further down the memory acquisition/formation/retrieval process there is again a mark of cognition when Otto desires to bring to his conscious mind the address of the museum. The intentionality of the process of storing MoMA's address for later recall remains throughout the acquisition and retrieval process. The words in the notebook are an engram of a different form; the end result of an encoding process of a different kind. The information on the page, as an engram is mentally inert until such time that another mental process (in this case the conscious desire to bring to mind the address of the museum) caused it to be brought into a working memory space. The same could be said of those parts of the brain responsible for storing a memory. Looking at the neurons themselves will not reveal any sign of intentionality. It is just a physical state as are the words on the paper. Think of this kind of dispositional mental state like a book that no one in the world is currently reading. The content is all there but until a reader picks up the book and, through reading, puts the ink on the page into her thoughts the book is powerless to do anything. The original content is not found in the notebook any more than it

would be found in a biological engram of a dispositional memory. The original content is found in the mental process as whole, i.e. - within *Otto*. Without context and the right means of understanding, the lines written in the notebook are meaningless. Likewise, an engram in the brain is a seemingly random series of neural connections and chemical states. Looking just in the brain itself tells us nothing about what the information represents, yet the information is there.

This nuance is captured in the four above criteria for extended cognition (readily accessible, transparent in use, constantly available and previously endorsed with future consequences) specifically the fourth criterion: previously endorsed. Unlike a squiggle in a notebook, Otto's action of recording information in the notebook was *about* something. He was not just throwing ink at the page and waiting to see what came out. There is no apparent reason we should think intentionality is lost because the information, qua information, is recorded on paper rather than in biological memory since it functions for Otto in the same way. The recording of the information in the notebook, when recorded, was *about* remembering the location of the museum, referring to the notebook to retrieve the information was *about* bringing the information to Otto's conscious mind. The intentionality remains in Otto and the information is in the notebook. The intentionality and the information qua information are separate cognitive processes of the same cognitive system. The fact that these parts are in different locations is irrelevant.

If Adams and Aizawa are correct this "aboutness" is lost when the memory is copied from biological memory to the notebook. But is this always the case? Certainly language, even in the hands of the most skilled writers, is inadequate to capture all the nuance and phenomenal

experience of a remembered event. But not all memories are focused on the experience of events. Some memories are just about facts.

Semantic vs. Episodic Memory

If I were asked to relate some of my memories of my childhood home I would tell stories of events that happened under that roof. I would relay my experiences of home. However, not all of my memories are of experiences. For example, I remember the address, where the rooms are located, and the color of the walls. Some of my memories are about events and others create a map.

Not all memories are of the same kind. There is also a difference between the memories accessed to use our native language to generate speech, and those involved in recognizing familiar faces. The memories which Otto stores in his notebook are of a particular kind. Not just any biologically held memory can translate to a written version without loss of fidelity or usefulness. The distinction in kinds of memories germane to this dissertation is found between what Endel Tulving calls semantic and episodic memories. Tulving defines each category of memory thusly:

> Episodic memory receives and stores information about temporally dated episodes or events, and temporal-spatial relations among these events. A perceptual event can be stored in the episodic system solely in terms of its perceptible properties or attributes, and it is always stored in terms of its autobiographical reference to already existing contents of the episodic memory store. (Tulving, 1984)

> Semantic memory is necessary for the use of language. It is a mental thesaurus, organized knowledge a person possesses about words and other verbal symbols, their meaning and referents, about relations among them, and about rules, formulas, and algorithms for the manipulations of these symbols, concepts and relations. (Tulving, 1983)

Adding a few pages later:

Information stored in the semantic memory system represents objects—general and specific, living and dead, past and present, simple and complex—concepts, relations, quantities, events, facts, propositions, and so on, detached from autobiographical reference. (Tulving, 1983)

Clark and Chalmers fail to distinguish between these two kinds of memories and I believe this leads to some of the confusion or skepticism behind the extended mind thesis. I will argue that the extended mind, as it pertains to memory, is capable of easily supporting *some* semantic memories but not (at least not, as easily) episodic memories. Once this distinction is made I believe the extended mind thesis will be on stronger ground. My dog, Houdini (AKA 'Doodle'), can serve as an excellent example for making this distinction[62].

It is sometimes said that dogs have no (long term) memories yet they seem to behave as if they remember plenty, like how to respond to commands or getting excited when an owner comes home from a long trip (or even a short trip since dogs have what seems to be a terrible sense of time). This is not a contradiction, but an equivocation based on a failure to differentiate between different kinds of memories.

Doodle gets very excited when I approach a particular cabinet drawer in my living room. This is where his treats are stored. Why would he get excited each and every time I do this if he has no memory of me doing this in the past? In truth, he does have memory of this – a semantic memory, not an episodic memory.

Through time and conditioning Houdini has developed a semantic memory, something like a heuristic which tells him "Michael+approaching+drawer = I+get+something+good"

[62] Since this analogy was drafted Houdini has passed away from a rare and unforeseeable autoimmune disorder. It feels like a betrayal to replace this analogy with another. May Doodle live on, even if as an extended memory of his loving owner and best buddy.

(reduce this to any level of description you like).[63] However, despite the conditioning he may not remember any particular instance of receiving a treat. Likewise, Houdini, being a rescue dog, demonstrates behaviors characteristic of a dog which has been abused (cowering when approached for example). His behavior comes from a mental map of having been abused, though he may not remember any particular instance of abuse. The same goes for human memory as well. I remember how to sing "Happy Birthday" yet do not remember either having been taught the song nor learning it through exposure.

Though I think the distinction between episodic and sematic memories is clear, I do think there is some possibility for confusion within the category 'semantic memory' itself. There are important differences between the kinds of information Tulving titles collectively 'semantic'. The mental qualities and activities required for the production of spoken language seem very different from those involved with the recall of brute facts or solving simple arithmetic. All of these are rightfully called semantic given that part of Tulving's definition of semantic memory is all memory which is not episodic (Tulving, 1972). But I believe a further category needs to be added to save the trouble of constantly qualifying what kind of semantic memory we are dealing with. I purpose adding a sub category of semantic memory and calling it 'isomorphic memory'[64] which deals with matters of fact learned or discovered by the one who remembers.

---

[63] There is a kind of perennial problem in discussing contents of minds and all things mental in general. To talk about the mental we must use the one tool at our disposal, language. It is an open question whether or not humans use a "language of thought" or "think in words" but certainly nonlinguistic animals and pre-linguistic humans do not. So, are we committing some kind of fallacy or act of hypocrisy by talking about Houdini's mental life in terms of words? I don't think so. Searle talks about his dog and his dog bowl. Though dogs don't think in language we can nonetheless exhaustively describe the dog's thoughts with language. I will address this connection between minds and talking about mind in chapter 4.

[64] This label was suggested by a mathematician friend Joe Kelly. In mathematics 'isomorphic' refers to having similar forms or relations. I will also refer to 'isomorphic information' and 'isomorphic content' meaning information which is the same across media.

The memories of the home I grew up in can be divided into three kinds. There are the episodic memories of specific events like family visiting after my grandparents died or amorous encounters with my high school girlfriend. There are also semantic memories, such as those of the layout of the house and the position of the furniture, which allowed me to navigate in the dark without bumping my shins or stubbing my toes, which I do not specifically remember learning but nonetheless are things I learned at some point (or more exactly learned over time)[65]. There are also isomorphic memories which are brute facts about the house—the room at the end of the hall is the smallest or the laundry room was the first door on your right as you walked down the hall from the living room—which would be true whether I ever lived in this house or had ever even visited.[66]

This distinction plays an important role in the extended mind, specifically in the primary example of Otto's notebook. Clark and Chalmers have given us a general idea of what is in the notebook, but only one specific example – MoMA's address. But from the way in which Clark and Chalmers describe the notebook we might assume that if the entries are memories then they are isomorphic memories.[67] The address alone tells us nothing of the love Otto feels for MoMA nor, much about what kind of thing MoMA is. The address is just a fact. The statement in the notebook of the museum's address is just as true in the notebook as it would be in a phonebook, smart phone or Otto's (or anyone else's) head. It is telling, however, that the address is there.

---

[65] Or even more exactly had to relearn over time. My mother seems to have a compulsion for rearranging furniture.

[66] In a sense, semantic memories such as "you cook in a kitchen and wash in the bathroom" are also brute facts. These examples are not chosen because of their factual nature but rather because they are things one has learned but nonetheless cannot remember having learned them. There will be some overlap between isomorphic and other semantic memories.

[67] Perhaps this is not what they meant, but I will move forward as if this is the case. Regardless, isomorphic memories are what I will argue are most easily and likely to extend.

The fact that Otto wrote the address down signifies that it is of some importance to him and presumably there is some consequence for Otto in it being there. Perhaps because he loves the museum and wishes to visit often or perhaps because it holds unpleasant memories and he does not want to go there again. We would be reliant on context to decide which of these two (or the many shades in-between) it is.

The fact that something is written in Otto's notebook is not enough to rightfully call it a part of his mind, a memory, mental state or what have you. There is a difference between a diary entry recounting the events of the day in an attempt to capture the feelings felt in those moments versus keeping record of brute facts. This is the difference between the mere fact of the address 1600 Pennsylvania Ave. and the episodic memories one may have from time spent touring the White House or the reverence or contempt one may feel for its current inhabitant. The mere fact of the building's address can be fully and completely accounted for in both biological memory and on paper equally well. A solely biological substrate is not required for the cognitive process of memory formation for information of this type. There are cognitive events and states which are not capable of surviving the transition from internal to external storage. So long as we do not try to endow the extended parts of the mind with phenomenal capacities we associate with the human brain, the extended mind theory remains a viable theory of mind.

The isomorphic/semantic/episodic memory distinction is much like the syntax/semantic distinction in information processes. The distinction between syntactical and semantic is a distinction in the kinds of meaning cognitive information contains. In describing the work performed by a syntactical system I will explain (roughly) how information flows through a modern computing system.

Syntactical systems are those concerned with manipulation symbols according to rules bereft of meaning (to that system). In a computer's calculator program the system looks something like this: if I press the '5' key on a keyboard an electrical circuit is completed and a series of logical gates open in the pattern 'OFF-OFF-ON-ON-OFF-ON-OFF-ON' (meaning electrical current flows through gates 3, 4, 6 and 8 only)[68] and the Arabic number 5 is displayed on the monitor. If I then press '+' (opening gate 3 only) the previous gate configuration is stored in a data accumulator for further use and the '+' is displayed. When '7' is pressed the following gate combination is opened 'OFF-OFF-ON-ON-OFF-ON-ON-ON' and the Arabic numeral '7' is output to the screen. When I hit '=' (opening gates 3 through 6 and 8) the combination of logic gates stored in the accumulator and the set opened when '7' was pressed are run through an algorithm to determine a new number for output. The algorithm opens the outgoing gate sequence 'OFF-OFF-ON-ON-OFF-OFF-OFF-ON-OFF-OFF-ON-ON-OFF-OFF-ON-OFF' and '12' is output to the display.

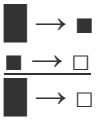
A computer performs this same kind of process for every action it takes. If a person is playing chess against a computer what has changed? The interface and the displayed result are different but internally a computer calculates a chess move in the same way it does arithmetic, only with more variables and more complicated algorithms. To a computer doing basic math, displaying a YouTube video, processing the text of this dissertation and playing chess "look" exactly the same. There is no meaning and no 'what-it-is-likeness' to its operations. The computer follows (sometimes very elaborate) instructions so we are justified to call it intelligent in some sense of the word, but also can, rightfully, be called mindless (or 'dumb) in the strictest

[68] This is not a necessary configuration of logic gates and, in fact, there are others found used in different types of hardware. This just happens to be the most common configuration and was arbitrarily decided upon when digital computers were first being programmed.

sense as these same computations can, in principle, be carried out by any universal Turing Machine. It is because of this mindless adherence to instruction and lack of meaning that we can argue that computers don't actually play chess (Haugeland, 1985) insofar as we want to say chess is *about* winning, or having a good time or challenging one's self or simply playing a game.

For all of the impressive feats our mind is capable of there is still, at some level and at some times, mere information processing going on. When we do addition in our heads there is both an intentional and syntactical aspect to the problem solving. On the one hand, we are trying to solve a problem. The activity of adding is *about* finding a sum for which we presumably have some other use. But the action of adding qua adding is just the manipulation of symbols.[69] As adults we can answer the question what is 17+12 with little difficulty (and indeed some problems such as 2+2 are perhaps closer to a reflex than a calculation), but I remember having trouble with simple multiplication as a child. If I had a homework assignment to calculate some multiplication problems my actions would be on one hand *about* solving the problems so I could finish my homework and thus be free to watch cartoons, but the actual process of solving the problems themselves would be the implementation of an algorithm (which probably involved counting on my fingers or shifting crayons into one pile or another and counting to get the

[69] If the use of numbers raises red flags for the reader (because the numerals 'mean' something) consider instead the solving of a logic problem which uses empty variables, such as this syllogism:



result). Within the semantics of wanting to get my homework done was a syntactical process making the end result possible.

The difference between me doing addition and a computer "doing addition" is in the reason for the end result. In my case the end result was the freedom to watch cartoons but for the computer it was an algorithmic move toward a certain machine state. Desire, want, or what have you, is the semantic element present in minds but absent in computers. Syntax is still present in the mind, it is just bookended by semantics on either side. There is a desire to solve the problem (semantic), the solving of the problem (syntactic) and the satisfaction of having solved the problem (semantic again). For me there was the semantic desire to solve the problem, the syntactic steps in solving the problem and the sematic satisfaction of getting to watch He-Man once the (hopefully) correct answer was reached.

In post dementia Otto we can also see both the presence of semantics and syntax in his extended mind. Otto does not go looking for MoMA's address for no reason. The recall of this extended memory is precipitated by an intentional, non-derived, desire to know the location of the museum. The content of the memory of the address is not semantic, it is just a fact (engramed as an isomorphic memory); the Museum of Modern Art is located at 11 W. 53rd St. in Manhattan. It does not matter if the memory is biological or extended. In both places the address is a mere fact with no need of intentionality in and of itself. The causal power of the information in the notebook creates the same functional intentional state within Otto after the information is retrieved. As such, there is just as much original content in the extended mind as there is in the unequivocal mind in this case.

The Coupling/Constitution Fallacy

Adams and Aizawa accuse Clark and Chalmers of making an elementary mistake in the classification of things: confusing the mere coupling of two things with the constitution of a single thing. Adams and Aizawa claim "it simply does not follow from the fact that process X is in some way causally connected to a cognitive process that X is thereby part of that cognitive processes" adding "one cannot simply move from the observation of a causal dependency between cognition and the body and the environment to the conclusion that cognition extended into the body and environment." (Adams and Aizawa, 2010).

An analogy used by Adams and Aizawa is bowling. The verb 'to bowl', in the context of the game, refers to the action of throwing a heavy ball, underhand, down an alley toward a set of pins at the far end. During the ball's journey it will be affected by the alley via friction and the alley's subtle topography, changing the speed, spin and thus direction of the ball. These changes in the direction and movement of the ball are causal and may affect the outcome of the throw but this does not mean that the effects of the alley are part of what it is 'to bowl'. One bowls by throwing the ball in the prescribed manner; as Adams and Aizawa say "there is no extended bowling" (Adams and Aizawa, 2010).

Bowling is a system that requires an alley, pins and a ball to come off. The alley plays a causal role in how the ball ultimately interacts (come into contact) with the pins.[70] Other factors such as the relative humidity, air flow and cleanliness of the approach also affect the outcome. But just because some external entities are required for an activity and other factors play a role

[70] For example, how heavy the wax on the alley and how it wears over time change the speed and spin of the ball. These effects are small enough that novice bowlers won't necessarily notice a difference, but to a practiced professional they are the difference between a great or so-so score.

does not mean these things together with the actual throwing of the ball constitute a single extended activity which is “to bowl”.

Adams and Aizawa admit that some cognitive activities may require the use of external tools (nearly everyone would have to write down a string of 47 numbers in order to recall them later) but it does not follow that these tools are themselves cognitive or part of the process of cognition; causal, for sure, but not necessarily cognitive.

The general claim is that just because two things (in the above analogy the act of bowling and the effects of the alley) are coupled in a causal relationship does not mean all parts of this coupling constitute a single entity. Some interactions are both causal and constitutive. Neurons in my brain working together to solve a math problem are certainly both causally connected and constitutive of a cognitive process, but Adams and Aizawa believe the bowling analogy shows this need not be the case with all causal relationships. Just because Otto’s notebook plays a role in the cognitive process of remembering where MoMA is located does not mean the notebook and Otto’s brain constitute a single thing, an (extended) mind. Nor does it mean that the coupled system of the notebook and the brain are together a new thing distinct from its individual parts. Of course, a coupled relationship does not always mean a constitutive relationship (if I give myself an aspirin for a headache I am not now a drug delivery device or an aspirin storage device) but I argue this depends entirely on the level of analysis which is appropriate to the phenomenon in question.

Just as Adams and Aizawa have accused Clark and Chalmers of making an elementary mistake as regarding kinds, so to I will accuse Adams and Aizawa of committing a different elementary fallacy: a category mistake.

A Category Mistake

Extended bowling seems a laughable notion so long as we restrict our thinking of 'bowling' as a mere verb describing the action of the individual. The swing of an arm and the consequent projection of a ball is a closed (i.e. – non-extended) system if ever there was one. If we view bowling (the noun referring to the game) as a process or system we see things differently. The effects the alley will have on the direction, spin and velocity of the ball are not entirely known before the game has progressed a few frames. After throwing a few balls down the alley an experienced bowler will be able to alter his throw to compensate for these small physical effects. There exists a feedback loop between the bowler and the alley. "Bowling" is a process in which some factors involved in the outcome are indeed things physically apart from the bowler. At this level of analysis the process of bowling extends past the bowler and encompasses external factors. Not only does talk about 'extended bowling' make sense when viewing bowling in this way it is necessary to fully capture what is happening. The tension between the different levels of analysis come from the fact that any *complete* description of the system or process of bowling must include the alley and its effects, however, this is not what we most often mean when we talk about bowling, therefore, it is not in many or most cases, an appropriate level of analysis of the verb 'to bowl'. Bowling may be a closed activity, limited to the person preforming the act, yet bowling is also an extended process which, as it unfolds, encompasses many factors beyond the actions of the person who initiated it. Perhaps all processes are extended in this way once we consider everything going in within a process. The shape of a highway, straight or winding, has an effect on how people drive. The invention of a new medical technology changes how we treat disease. It might be strange to talk about extended bowling, driving or medicine, but it is undeniable there are external forces and causes

at play in all of these activities, perhaps so deeply entwined with these practices the causes or forces may well be constitutive of the processes.

Once we start talking about the mind, however, people start to get nervous. For so long our own knowledge of our own mind has been so certain and so completely private that talk of different levels of analysis seemed at least odd, if not truly impossible. The extended mind makes us face questions such as what kind of thing is the mind and consequently what is the right level of analysis for talking about minds?

What I propose to do is offer a plausible description of the mind which answers Adams and Aizawa's objections and keeps open the door for extended cognition. Adams and Aizawa have made an assumption about the mind which they have not adequately proven. They assume there is a difference between a cognitive system and a cognitive process; that the mind is a thing separate from either the system upon which occurs, functions, supervenes or what have you, or the activity that it performs. I argue that a cognitive system and a cognitive process are in fact the same thing. A cognitive system is a system made of a set of processes, not a thing in and of itself apart from those processes. In what follows I will consider the different ways in which things can be coupled. Sometimes things can be coupled in such a way that each retains its own identity while at other times the coupling or decoupling result in the obliteration of the objects' existences as they once were. I will then discuss the idea that a cognitive process and a cognitive system are one and the same thing, thus defining, at least roughly, what kind of thing the mind is. Lastly, I will look at the different possibilities as to what happens when minds are coupled with the world in the form of other minds or non-cognitive entities.

Different Kinds of Coupling

Let's say I have two pieces of Silly Putty, one in each hand. I am in possession of two distinct things.[71] But if I squish them together and knead them for a bit I am in possession of the same mass but now in the form of a single thing. The individual masses of Silly Putty cease to exist in that the particular token assemblages are gone, despite the mass remaining constant. If I break it apart I again have two distinct things, made of the same matter yet numerically different from the two things I had before, in that each is, in all reasonable probability, going to be a different token assemblage of matter. In this case the coupling and decoupling of the object resulted in numerically different things. But this is not the case with all instances of coupling and decoupling. Some things can be coupled, decoupled and coupled again while still retaining their own numeric identity as well as participating in the identity of something else.

A horse fitted with saddle does not constitute a single thing by means of the obliteration of each in the creation of the new. It is not the case that once a horse is wearing a saddle the horse and saddle cease to be and a (let's call it) horsaddle now exists in their place. However, this may depend on the level of description we are using when discussing the horse and the saddle. At the level of thing qua thing (the level we most often use when discussing mere objects in polite conversation) the horse and the saddle remain separate entities regardless of the configuration as they are easily separated into discrete objects. The horse and saddle remain different things regardless of whether the horse is wearing the saddle, the saddle is on the storage rack or the saddle is on the leather worker's bench. But, from a different perspective something new has been produced when we couple horse and saddle. The functionality of the two entities

[71] Overthinking what is meant by 'thing' in this case will not be at all helpful. Whether an apple is a single thing or two things (2 half apples) or a collection of fruit, seeds, stem and skin will only add to the number of entities we need to consider, not change the argument.

has changed. This is a new *system* if not a new thing. Soldiers have used horses with saddles for millennia, as there is a functional advantage over bareback riding. The addition of stirrups was yet a further functional advantage. But each addition of technology did not change the kind of thing the rider or the horse or the saddle was, but rather only increased the abilities of each. The saddle and the stirrups did not create something different form riding; they merely improved the range of options available to the rider.

Gilbert Ryle, in response to Descartes, and what Ryle calls the "Official Doctrine" (Ryle, 1949) argued that Cartesian substance dualism rested on the mistake that the mind was a thing (i.e. – part, of a human being just as is an arm or pancreas). Though it is perfectly fine to talk of someone's "keen mind" (much as we would refer to that person's 'strong arms') we must be careful not to mistake the mind for a thing in the same sense that a physical object is a thing. Whether the mind is a system or function or property or what have you, it is still a thing but a distinct kind of thing from those things which make up the physical half of a dualist entity. As Ryle says, a parade is a thing but it is not a thing I can show you apart from the parts that make it up as the event takes place.

Otto and his notebook are not, as Clark and Chalmers would describe it, a single unified, physical entity. While Otto carries the notebook he is not a different thing than he was before. Otto is still Otto and the notebook is still a notebook. However, the functionality of Otto's mind has changed. Otto is able to perform cognitive tasks which were not possible were it not for the keeping of the notebook. Adams and Aizawa are, no doubt, very familiar with Ryle's category mistake I do not wish to accuse them of making such an elementary error, per se. The mind is not a physical thing like a pancreas and they are well aware of this. The mind is a thing only in

that ‘mind’ is a noun about which we can say things. Unicorns do not exist yet we can nonetheless talk about them perfectly coherently.

I believe the distinction between a cognitive system and a cognitive process is a more complex kind of category mistake. Since it is about reshaping the boundaries of the mind (and perhaps even the self) the extended mind thesis calls into question the shape, form and location of a cognitive system but I believe what a mind is remains the same in both the orthodox and extended forms. I do not believe it is possible to locate a cognitive system apart from cognitive processes. One necessitates the presence of the other. A cognitive system simply is whatever supports a cognitive process. I may be accused of circular reasoning here if the reader interprets the above claim as “a cognitive system is a system capable of cognitive processes and a cognitive process is what occurs in a cognitive system.” This is a fair interpretation, but it is not what I intend. This circle only circles so long as we differentiate between the system and the process. I do not endorse such a distinction. A cognitive system *is* a set of cognitive processes. You cannot have one without the other. Something is a cognitive system just in case it engages in cognitive processes. It is no more circular than any other non-tautological bi-conditional statement. What makes something a cognitive system is the fact that it engages in a cognitive process. Conversely something is a cognitive process by virtue of being the function of a cognitive system. If there is a cognitive process, there must be a cognitive system on which it is occurring and if a system engages in cognitive processes it is a cognitive system. To talk about a cognitive system and a cognitive process is in some cases a redundancy, as you cannot have one without the other, yet it is entirely possible to discuss them separately as one refers to the medium on which something is happening and the other is referring to the activity taking place within that medium. The category mistake made by Adams and Aizawa is to force a wedge

between the two and treat them as separate entities. There are no necessary and sufficient conditions for a cognitive system other than the ability to support cognitive processes. We do face the difficult prospect of identifying cognitive processes/systems. A conscious adult human is certainly a cognizer, and thus, a cognitive system but what about a thermostat? Does the ability to identify temperature difference and alter internal states according to a program or other machine state constitute something cognitive?

Looking back to Parfit's divided, transplanted brain we see the possibility of a cognitive system (two halves of the same brain separated by physical distance) as possibly supporting a single cognitive process. But we need not go to such extremes to see the impossibility of separating the system for the processes. Playing off one of Ryle's example, consider the thing that is Michigan State University. We can rightfully talk about the thing that is MSU, but we can also talk about the processes and parts that make up the university. The thing constituting the university includes departments, administration, faculty and students, buildings, books, animals, laboratories, roads, the Dairy Store, sporting venues, libraries, etc. These entities all have relationships to one another and the interactions between them constitute processes that define the nature of a university. Michigan State University is a complex, compound thing we have a hard time defining, but nonetheless are perfectly able to talk about. We can strip away many parts of the university and still be left with something that functions as a university (teaching students, doing research etc.). In fact, MSU existed long before most of the parts that constitute it today.

Cognitive systems are very much like this. As discussed earlier, a human brain (which I assume Adams and Aizawa would agree is a cognitive system, if anything is) can function with a great portion of its mass missing, e.g. - as a result of a Hemispherectomy. It is not the case that a

Hemispherectomy only removes superfluous portions of the cognitive systems. If we were to perform a Hemispherectomy on a healthy brain we could have, in theory, just as easily removed either half of the brain and achieve different, yet successful, results. There appears to be no one part, or vital combination of parts that constitute this cognitive system. Though the amygdala is necessary for memory formation and emotional response, you can live without it and you are still a cognitive system, albeit one with a different set of capabilities. Broca's area is a key component in speech production, but damage to this area of the brain does not mean the person is no longer cognitive, only that he will be unable to speak or will have a great deal of trouble speaking. A cognitive system is just a system capable of sustaining a cognitive process, no matter what the token examples of a cognitive process turn out to be. The extended mind thesis merely argues that the substrates which make cognitive processes possible may change but the processes are nonetheless cognitive. If Otto's brain becomes damaged and he replaces a few neurons with a computer chip he is still a cognitive system engaging in cognitive processes. The same argument is made for replacing damaged brain matter with a notebook. The particular parts of the cognitive system may have changed, but the cognitive processes remain the same. The only difference is that the 'chip' is no longer in the head in the second case.

So What is Cognitive After All?

Whatever the means of identification of a cognitive system or process may turn out to be, we can know that if we find one we have necessarily found the other. This does still leave the problem of identifying where the boundaries of the mind lie. There is a world of difference between a well-organized notebook and standing in the middle of a research library. Thankfully this issue has already been, in part, addressed in the form of the four criteria for the extended mind. By these criteria the notebook can be/is part of a cognitive process, but the library is not.

What is missing here is a justification of when something, especially something external, becomes part of a cognitive system. Why does this particular notebook count and why not any notebook on the shelf at Staples? Why does or doesn't the pen used to make the notes count?

We don't face these temporal claims about the human brain qua cognitive system since the entire brain is firing on all cylinders, so to speak, most of the time.[72] As the brain develops as we pass through infancy, childhood, adolescence and adulthood certain parts come online at different times, but insofar as my brain functions after I have reached adulthood, it is all part of a cognitive system all of the time. Even if the neurons or brain structures related to the storage of a dispositional memory are not fully active at any particular moment, they are nonetheless still doing cognitive work as they store the memory and await and search for cues that it is time to bring the memory in conscious space. A notebook on the shelf at Staples has not yet been drafted into a cognitive process. It is no more part of a cognitive system than it was when it was still a tree awaiting logging or recycled pulp waiting to become new paper. The notebook becomes part of a cognitive process when a cognitive system conscripts it into service. It remains part of that cognitive system until either the system, implicitly or tacitly, discharges it from service (implicitly by casting the notebook aside or tacitly by social convention of not requiring someone to remember a three-digit exchange for more than a few moments) or is destroyed (Otto burns the notebook, or it is destroyed in a paper shredder mishap).[73]

We are still faced with a problem of determining when the adding in of a cognitive aid is conscription into the cognitive system and when it is a mere tool. Standing in the middle of a

[72] The idea that human beings only use 10% (or whatever percentage you'd like) of their brains is a myth, the origins of which are unclear, but has been resoundingly debunked. (Boyd, 2008 and others too numerous to mention.)

[73] Damaged, nonfunctioning or absent neurons are no longer part of a cognitive system either.

library has been established an inadequate integration of the library into one's cognitive process, but what about while actively engaged with the internet? Is the internet, or at least the webpage one is currently viewing, part of someone's cognitive process?

To answer this question let's consider one way in which this indeed would count as mind extension. Rather than a notebook Otto decides to organize his external information on a webpage. Otto still enters and retrieves the information for the same reasons he did with the notebook but instead of reading the words of a piece of paper he sees it on a screen. This qualifies as mind extension (should the criteria be met). It is not significantly different than choosing a lined notebook over one with unruled pages. There may be some technical differences, but the same isomorphic information is recorded, displayed and recalled.

One question remains through this entire discussion. What makes some processes cognitive and others not? I do believe Adams and Aizawa were right to point out the need for a 'mark of cognition'. They pointed to the presence of non-derived content as such a mark. However, this leads to another question about which there is no agreement: is there a such thing as non-derived content? Until this debate is settled there is going to be disagreement as to whether or not this is an appropriate mark. I granted, for the sake of argument, such content exists and has shown where within the extended mind non-derived content can be found: within the brain of the person whose mind had extended. In essence, this amounts to a claim that there can be no extended cognition without a (functioning) brain being involved in the processes. (Perhaps, however, there is some version of the extended mind thesis that can get away from this requirement.) Regardless, all the cases that have been, and will be, discussed in this dissertation are ones in which a brain is at work – be it the person whose mind has been extended or the person who is coming into contact with an extended cognitive apparatus.

This being the case, Adams and Aizawa's insistence on a mark of cognition is misplaced here. In cases where Clark and Chalmers and other proponents of the extended mind claim extended cognition is occurring, there is one thing Adams and Aizawa and opponents agree on: cognition (extended or not) is in fact occurring. When Otto checks his notebook for MoMA's address he is engaging in a cognitive function. All we are arguing over is whether or not the notebook is part of that cognitive system or just playing a causal role in the cognition going on in Otto's head. We have had a benchmark for determining the presence of cognition in front of us the whole time; if a human mind is involved, the process in question is a cognitive process. What we need to determine is whether or not the causal participants in a cognitive process are themselves cognitive (thus constituent parts of cognition) or merely causal (playing a role in cognition but not themselves cognitive).

In some cases this question has an obvious answer. If the mind in question is engaging in speech production then Broca's area of the brain is playing a causal role and is also part of the cognitive machinery. But Broca's area is part of the brain, the brain is made of organic matter and organic matter is made of molecules, atoms and ions. Are the particular oscillations of these ions themselves cognitive? Most likely not (Adams and Aizawa, 2010). The heart circulates blood, which feeds the brain the energy it needs to continue functioning. Is the heart cognitive? Extremely doubtful.[74]

[74] The hedging language 'most likely not' and 'doubtful' are here merely out of an acknowledgment of the theory that perhaps consciousness is a fundamental property of the universe we live in. The theory being that atoms are conscious, just not in a very interesting way. The consciousness we have is the result of the accumulated consciousness of the fundamental particles which make up or physical bodies. I am not sold on this idea, but I do find it a fascinating and at least plausible view of existence.

There are other cases which are not as straight forward. Nöe (2004) poses the question of whether or not the tongue is cognitive when engaged in the process of tasting wine if wine could not be tasted without it.[75]

> What is the causal substrate of the experience of the wine's flavor? Perhaps this substrate is only neural, but perhaps it is not. For example, perhaps the only way or the only biologically possible way to produce just the flavor sensations one enjoys when one sips a wine is by rolling a liquid across one's tongue. In that case, the liquid, the tongue, and the rolling action would be part of the physical substrate for the experience's occurrence. (Nöe, 2004)

The argument here being that in order to experience the taste of wine it is causally necessary for the tongue to be involved. Without the tongue there would be no taste. We are not talking about reading a description of the taste of a wine and arguing that the words being read constitute "extended tasting". The tongue is indispensable to the cognitive task of tasting, therefore the tongue is doing cognitive work.

But is it? By looking at other ways in which a person can come to have a sensation of taste we can see why the tongue may be excluded as a cognitive entity.

The tongue plays a causal role in the case discussed by Nöe. It is the means by which the chemical cues are converted into neural signals and thus transmitted to the brain.[76] One cannot taste wine by pouring it directly onto one's gray matter. Tasting takes place in the brain, and is experience *as if* it is occurring on the tongue.

---

[75] The tongue is not in fact necessary. Nöe, no doubt is aware of this fact but for the purposes of the thought experiment concerning the necessity of an external entity quite defensively assumed otherwise. This will be addressed shortly.

[76] Strictly speaking, the sensory nerves in the nose play a large causal role in taste as well. So much so that it can be argued that smell and taste are not actually different senses. For the sake of simplicity, we will ignore this gustatory fact.

By way of analogy consider the question: if a tree falls in the forest and no one is there to hear it, does it make a sound? This question is trotted out time and again as some kind of impregnable Buddhist koan, but the answer is quite simple: no, it does not make a sound. Yes, a pressure wave is produced (which we also call a sound wave), but sound as an experienced phenomenon is produced in the brain of the hearer. It is the result of the pressure wave hitting the ear drum and the brain interrupting the signal as sound, which is experienced as if it is happening in the ear. The surface of the sun is a violent place, home to nuclear explosions. If you could survive the heat and the blast waves, we would find the sun to be an incredibly loud place.[77] However, there are no things on the sun capable of hearing, so there is no sound, only sound waves. When we say 'sound' we are referring to the phenomenal experience, not the physical wave which proceeded it.

Regardless of the place in which taste 'happens' does it matter if the tongue is casually necessary for that sensation to occur? Is the fact that the tongue is a necessary part of the cognitive process enough to make it a cognitive part of the process? Adams and Aizawa say no.

> Why does the uniqueness of the cause of the perception provide any reason to think that the cause … is part of the physical substrate of the perceptual experience? Suppose that the only physically possible way in which to initiate a nuclear fusion reaction involving hydrogen would be through an explosive nuclear fission reaction involving ^{238}U.[78] Would that be any grounds for saying that the fission of ^{238}U is fusion of hydrogen? (Adams and Aizawa, 2010)

Here again Adams and Aizawa are demonstrating their instance that causal connection does not equal a constitutive relationship, and I think they have given a persuasive example of their point, in general. I think they are right that the fission is not part of the fusion. ^{238}U can, in theory, engage in fission in the absence of hydrogen and thus no fusion of hydrogen takes place. It is not

[77] Much louder than the 180-200 dB (mentioned in a footnote in chapter 1), which would kill a person.
[78] Uranium 238. The most common isotope of uranium found in nature.

part of the fusion of hydrogen insofar as that fusion is an event having a discrete spatial/temporal existence, but it is part of the fusion of hydrogen qua process. The process could not take part without it.

The mind is a system made up of processes which are cognitive by definition. Tasting is a cognitive process taking place in the brain. The tongue happens to be the means by which this cognitive process gets initiated. Limitations of the thought experiment aside, in truth the tongue is not necessary. What is important in the sensation of tasting is that the relevant areas of the brain receive and interrupt appropriate incoming signals and this can be accomplished through other means.

The father of a friend of mine suffered an industrial accident when he was a teenager. The after effect of the accident was that his taste buds were all but dead and not coming back. He has not had the experience of taste for decades, except for a mild sensation of heat from extremely spicy dishes.[79] He chooses his meals based on temperature (as in Fahrenheit) and texture. Cold and smooth dishes (like ice cream) or crunchy, room temperature snacks (like nuts and chips) are preferable. But, so far as anyone knows, the areas of his brain that would otherwise receive and interpret signals from the tongue escaped unscathed. As is done during exploratory, open skull brain surgery requiring mapping of the brain we can stimulate these areas with mild electrical current and give him the same sensation of taste without the tongue or food ever becoming involved. Though his taste buds are dead he could still experience taste as if they were functioning normally because the actual experience of taste takes place in the brain. We might call this *artificial* tasting, but it is still tasting nonetheless, and therefore cognitive.

[79] The kind of spicy that most people would find intolerable.

I argue that neither the medical device used to stimulate the brain into having sensation or the tongue are cognitive but merely causal. To do so we'll consider the four criteria for the extended mind.

- readily available
- easily accessible (transparent in use)
- a constant in one's life
- must have been endorsed in the past and there is some consequence to its future endorsement. (Clark and Chalmers, 1998 and Levy, 2007)

The tongue is certainly readily available, easily accessible and a constant in one's life. But do we endorse it? In the end we do not really *decide* if something tastes good or not. Certainly, we can choose to eat a food we do not find particularly appealing and acquire a taste for it (as I have done with coffee and scotch). But this is still not the same as choosing to think something tastes good (I cannot, try as I might, find eggs appetizing).

What makes Otto's notebook and his tongue cognitively different is the means by which he came to "endorse" what each had to tell him. Otto made a conscious decision to make something part of his extended mind, whereas his tongue has always been telling him what he likes and doesn't. This is not to say that what is extended must be explicitly endorsed. We can tacitly incorporate the external into our minds. Unusually interdependent couples have come to extend their minds into each other through habit and practice, not necessarily through explicit cooption.

There is also a difference in the kind of information found in the notebook and the kind of information which comes from tasting. When someone tastes wine (by drinking it or through

electrical stimulation) there is cognition going on to be sure, but is the information being relayed from the tongue to the brain the right kind of information? Is that information always going to be interpreted the same way? Otto can say "I will believe what the notebook tells me regarding the address of MoMA" and likely will do so if he believes he is the one who wrote that information down. By doing so, Otto has off loaded the cognitive task of remembering this one bit of isomorphic information. If he were to write in his notebook "I like the taste of strawberries" he might be led to believe that if he were to eat a strawberry in the future he would enjoy it, but having this in the notebook does not make this the case. Along with everything else going on with his degrading brain, maybe his tastes have changed. Taste is not isomorphic memory, or a memory of any kind. No two people experience taste quite the same way (we all like different things) and the sensation of taste cannot survive as is moves across media. A description of wine cannot convey taste information the same way the Yellow Pages can convey information about a shop's location. This requirement of intentional endorsement and isomorphism are not requirements for cognition in general. My love of pizza is not isomorphic nor intentionally created, but certainly such strong feelings are cognitive and possibly the result of non-derived mental states. It is a fact I have eaten pizza in the past and this is a fact that can be fully captured in the writing of the words "I have eaten pizza in the past." The memory of having eaten pizza is capturable by an extended cognitive apparatus, but the semantic experience of what it was like to eat that pizza is not. Both the memory of the event and the memory of what it was like are equally cognitive, however, they are not both extendable tokens of cognition. The requirements of endorsement and isomorphism only apply to extended aspects of extended cognition as their functionality remains the same whether extended or internal.

Extended Mind, Number of Minds

In the next chapter I will discuss the implications of the extended mind on proxy decision making and advance directives. These implications are based on the idea that a single mind can span more than one brain or non-neural substrate, thus the revelation of mental states can be accomplished through varied mediums. Consider, again, the phone number example. In order to lessen the cognitive burden of memory when in need of remembering a phone number we can turn to the person next to us and ask him to remember the three-digit exchange so we can focus on the four digit line number. When we do so we might ask the question: exactly how many minds are involved in this cognitive task? Arguments can be made that there are as few as one mind and as many as three. I will briefly explore each of these options but conclude that there are two minds at work, but not in the orthodox sense.

<u>There are 2 minds- The Orthodox View</u>

Under this view the phone number example is properly described as one mind being in need of remembering a phone number and another mind is serving as a cognitive *aid* in this task rather than actually part of it. There is no metaphysical funny business going on here. The case is not different in any meaningful way from a person using a pen and paper to remember a phone number.[80] There is the mind of the person who needs to remember the phone number and the mind of the person who is nice enough to help with the task. This is merely aided cognition, not extended cognition.

This view fails to account for the functional argument at the base of Clark and Chalmers work. Here, the second person's action of remembering the three-digit exchange is the work of a

[80] However, a proponent of the extended mind thesis may agree that this is no different than using a pen and paper, yet still claim both as examples of extended cognition in actions.

mind distinct from that of the person wishing to recall the whole phone number. However, if the cognitive act of remembering the three-digit exchange were to have taken place in the agent's own head we would consider it, without question, as a cognitive process belonging to that person.

Since this is the orthodox view and this dissertation is largely an argument against the orthodox view, I will not give further objection here but say only that this view has the number of minds right but doesn't describe the nature of those minds properly. A fuller explanation will follow in a few sections.

There is 1 mind

A cognitive task must belong to a single mind. The task in question here is one of remembering and each act of memory is the act of a single mind. Two minds can remember the same kind of thing, but each is a different instance of remembering taking place in a different, discreet mind. If the act of remembering the phone number spans two brains and only one mind can remember any token memory, then these brains must be components of the same mind.

Again, there are some elements here which I will agree to (e.g.-both brains are participating in the same mind), but the overall claim is too strong (that the brains are completely inseparable from that mind). The very idea that there is one, and only one, mind in this kind of extended cognition has at least one serious problem. Though we are talking about a single cognitive task (remembering the phone number) we have failed to take into account the fact that many mental tasks happen within our brains simultaneously, in parallel. It is not the case that the person asked to remember the exchange ceased to have a mental life of her own because of the undue cognitive burden of remembering. Just because I ask the person next to me to remember three numbers does not mean our minds have merged and we are no longer separate persons. It

seems no matter how entangled two person's minds may become there is still some separation of mental lives. For example, the spatial location of each individual will mean each will process visual stimuli in slightly different ways, even if they are looking at the same thing. Each brain is a warehouse of a different set of experiences. Each person will have their own likes, dislikes and opinions.

The extended mind thesis is not one about the merging of minds to create a super-mind or the triumph of one mind over another.[81] It is about distributing cognitive burden and utilizing unused or underused resources available to the mind which is attempting to extend. The person remembering the exchange can easily go about her day, engaging in cognitive tasks unrelated to this particular act of remembering and still, under the right circumstances be said to be part of another person's cognitive apparatus. The mind extends and the person tasked with remembering is holding the memory for two: the person who asked for the mental favor and she herself. This is one (numerically, not qualitatively) memory held by two persons and not memories held by one person.

There are 3 minds

This is closer to what I think is going on but still slightly off the mark. The three minds in question are: one belonging to each person and a third which is constituted by the cognitive process of remembering which is spread across the two brains.

As with my discussion of Parfit earlier in this chapter concerning the theoretical transplant of two hemispheres of the same brain into two different bodies, here there is a single cognitive process spanning two separate brains, neither of which alone is sufficient for the carrying out of that task. However, these cases are importantly different. With the Parfit

[81] Though perhaps these are conceptual possibilities.

example we started with something that was, without serious question, a mind. The accident victim-cum-brain donor had what all reasonable people would call a single mind. There is a history, continuity and robustness which is absent in the "2 brain, 1 phone number" memory scheme. The substrate which supports the memory of the phone number is too tenuous, isolated and task specific to be considered a process capable of constituting a mind in and of itself. In addition, if the mere act of recalling a short series of numbers upon command (which is the only cognitive task or ability constituting the third mind in this case) were all that was required to be said to have a mind we would not have need of the Turing Test nor any debate about what things have minds or don't, as any number of machines or non-human animals can be made or trained to carry out this task. Whatever a mind is, it is clearly more than something capable of memorizing three numbers.

There are 2 minds (redeux)

This is what I argue is the case. There are two minds. One is constituted solely by the cognitive tasks performed by the conscripted person, including the task of remembering the three-digit exchange and the other mind is constituted by the conscriptor's set of internal cognitive processes and the external process of remembering the three-digit exchange. This is the view I will defend here.

The cognitive task of remembering the three-digit exchange is remembered by two persons utilizing the same brain, the brain of the conscript. The conscriptor can call upon the drafted person's brain to recall the numbers, but the conscripted person may make use of these numbers in any other way she sees fit. Perhaps she thinks it would make a good PIN number or wishes to play the numbers in that evening's lottery.

Defending the redeux

Consider the divided and transplanted brain. Parfit gives us a plausible argument that the human brain can be extended in this fashion. If we assume an orthodox view that the mind is inextricably linked to, tied up with, caused by or made up of the brain, isn't it the case that this brain extension was also an extension of the mind of the donor person? If neural structures for memory A are located in one half of the brain and neural structures for memory B are located in the other half, then the mind is indeed extended, perhaps even over great distances. There is no need for the two halves of the brain to be located spatially near one another in order for the mental states to be the mental states of the same person. This situation seems at first to fail to meet the requirement for extended cognition laid out by Clark and Chalmers, and Levy, but the problems of ready and reliable access are not insurmountable here. Communication between the halves of the brain can be achieved through face-to-face contact but also over wired and wireless communication. Otto does not have perfect access to his notebook and, as previously argued, we should not require perfect access in any case. The degree to which the access is affected is merely the degree to which the extended mind can work quickly and effectively, not a question of whether it can work at all.

The divided and transplanted brain thought experiment is not a perfect analogy for the extended mind thesis as discussed in recent literature. In Parfit's imagined case we are starting with a single brain and physically separating it, whereas the extended mind thesis is about the mind working its own way into the world and possibly spanning different brains and making a home in a totally different neural substrate.

How can we bridge this gap in this dis-analogy? I see two means by which we might do so. The first is by comparing the act of remembering within a whole (i.e.- undivided) brain

versus remembering once the brain has been divided. The second has to do with a comparison between how information "moves" within *a* brain versus how it moves *between* brains.

If I am alone, without a means to write and need to remember a phone number I have no choice but to try my best to commit the number to my own biological memory. Once I choose to commit the number to memory the means by which the information is stored is out of my control. In the cognitive processes of memory formation the information is encoded in some neural/chemical structure or other; I play no conscious role in determining how this information is stored nor have any control over how it is retrieved (aside from times *when* I want to recall the information into my conscious mind).

Suppose if after I have committed the phone number to memory I have my brain divided and transplanted. Suppose further that the transplanted half is where the neural structures for memory of the phone number are located. So now the memory is found in a new body, call it body X. If the other half of the brain located in body Y (the same body as before the division) has need of this information it will have to solicit the information from the brain half located in body X. The desire for recall and the task of recalling are now an extended task. One body will have to ask the other for the information.

How does the situation change if we substitute the brain dividing and transplantation by simply asking the person next to me to remember the three-digit exchange while I commit my biological resources to memorizing the line number? I still have made the conscious choice/effort to have the number remembered. In neither case do I have control over exactly how the number is encoded in biological memory and, when the information is needed, must solicit the information from a neural substrate different from the substrate in which the desire for recall was formed. In both cases the cognitive task of recall was extended over two separate

neural substrates. The fact that in the transplant case the substrates were once joined as a single, larger substrate is functionally irrelevant.

Is the Mind Already Extended?

The mere fact that extended cognition is extended is a problem for some, believing that a mind or a person has a distinct location and we are doing something conceptually dubious when we start talking about a person's mind existing (residing) in places it has never been. How can my mind exist, in part, in a brain which is currently located in, say, Hong Kong, or any other place I have never been?

Neither the brain nor mind is a geometric point which resides in one and only one place at a given time. Nor is the brain or mind capable of instantaneous action. The mind, as a function, property or what have you of the brain covers a finite but not infinitely small spatial area and performs its work over time.

The details and implications of this fact were previously discussed as part of the Parfit's divided brain thought experiment, so I will not rehash the issue at present, but I will add that insofar as we think of the mind being where the brain is, we face a further difficulty in defining where the brain begins and ends. The brain is made of about 100 billion neurons. But neurons are found throughout the body connecting the brain to the eyes, ears, fingers and toes. The nervous system extends throughout the body and our conception of the brain as the "grey matter" in our skulls is an incomplete picture. There are physiological differences between neurons found in the brain and those reaching into our finger tips. But there are also differences between neurons in various parts of the brain. From an evolutionary and physiological analysis, the eyes (more precisely the retina) are actually *part of* the central nervous system, made of brain tissue, not just an organ sending information along to the brain.

What does this have to do with the extended mind? The extended mind is nothing more than a stretching of these delays and gaps. Daniel Dennett's much discussed short story and thought experiment titled "Where am I?" pushes this idea even further. Imagine the brain is removed and placed in a vat and instead of neuronal signals moving information back and forth we have radio signals. Under such circumstances we could have a body in Michigan controlled by a brain in Kentucky. There would be a time lag between the sending and receiving of data, but this should not bother us, as there already exist delays in our skulls.[82] By separating the brain into different bodies we have only slowed the connection and altered the means of information transmission. But the speed is not what is relevant nor, is the method of transmission, so long as the information is of a kind that can survive the transfer. In a sense our minds are already extended within our heads.

Conclusion

There is no doubt that human beings use their environment and tools in it to lessen cognitive burden and expand cognitive abilities. Is this just functionalism or something more? The argument for the extended mind put forward above comes down to this: the extended mind is possible because minds, being intimately connected to brains, are by their nature extended. The debate over the extended mind thesis is really a debate about whether or not, or in what ways, the mind can extended further than the brain.

The analogies between neural processes and externalized substitutes or stand-ins for those processes are, in at least some cases, so much like neural processes we have little or no

[82] There are different effects on function as the delay increases. An 80-millisecond delay is virtually unnoticeable but a 1 second delay (the time it takes a radio signal to travel about ¾ of the way to the moon) would be quite noticeable.

principled reason to not properly label what is external as cognitive. To do otherwise is a bias in favor of our cranial activities.

Some concern over the extended mind thesis is based in the orthodox Cartesian idea that the mind is a closed, first person, privileged system. However, I have limited the scope of the extended mind in such a way that the specialness of persons should not come into question. Only a limited subset of cognitive processes, those concerned with isomorphic information, can extend. In short, facts are facts, regardless of where the information that constitutes those facts is found.

Next, I will begin to discuss the implications the extended mind thesis has for medical proxy decision making and advanced directives. I will first turn my attention to the problems of personhood.

Chapter 3- Problems of Personhood

There are two overarching issues regarding personhood and advance directives. On the one hand some argue (Dresser, 1986, Buford, 2008) that dementia can make it the case that a patient (under commonly accepted criteria of personal identity regarding continuation of some kind of mental characteristic) is no longer the same person as she who drafted the directive at some time in the past. On the other hand, some patients are in such states (PVS or the most severe, advanced stages of dementia) we can no longer correctly refer to them as persons at all (using the same criteria). These changes in person raise the question as to whether or not an advance directive applies, since the person who drafted the directive no longer exists; either she has become a numerically, metaphysically different person, or ceased to be a person at all. There are some similarities between these two kinds of cases (in both cases the person who drafted the directive no longer exists) but there is a crucial difference: in the first type of case there exists some person or other while in the second case there is no person at all. This means the extended mind theory will apply differently and lead to different results in each kind of case.

Addressing Dresser

One of the more persistent objections to adherence to advance directives is that the person who wrote the directive and the patient to whom it now applies are, in fact, different person based on the absence of continuity of mental characteristics. It is, all things being equal, wrong for one person to determine the fate of a different person without that determinee's consent in the matter. Some (Dresser, 1986, Buchanan and Brock, 1990, Brock, 1996, May, 1997, Tollefsen, 1999) have raised the question as to whether or not such application of advance directives on demented, yet once competent individuals, amounts to a form of slavery or other tyranny. (Buchanan, 1998, Tollefsen, 1999, Singer, 2003)

The task at hand here is twofold: 1) determine if a demented patient is or is not the numerically same person as the drafter of the directive and 2) determine if the advance directive should be applied. I will argue that a demented patient is the same person over time, no matter how pleasantly or severely demented the patient becomes, so long as there exists some degree of the patient's previous cognitive processes in the form of semantic memory. Only in a very small number of cases should we say a person is a different person as the one with whom she shares a bodily history and most likely a change of person will only manifest as a loss of all claim to personhood. To make my case I will call upon the work of philosophers separated by centuries, Thomas Reid and Derek Parfit.

Once I have been able to establish this claim, even if not to everyone's satisfaction, I will apply the extended mind thesis to these cases with two goals in mind: 1) the extended mind thesis adds strength to the claim that these persons remain the same persons over time and 2) to explain how the kind of dementia suffered by the patient (are they pleasantly demented or otherwise) and the dictates of the directive interact. In some combinations the directive should be followed, while in others not.

This sameness of person is not the end of the story and alone does not alleviate the concerns of Dresser et al. This is merely the starting point for the analysis of the implications of the extended mind in these cases. Cases where the patient has lost all claim to personhood (i.e. - death, PVS or most severe cases of dementia where no signs of meaningful mental life are present) are different and will be considered separately.

Memory Theory, Transitivity and Sameness of Persons

Thus far in this dissertation I have spoken about personal identity as if it is unquestionably a matter of psychological criteria, however, there are theories of personal identity

which argue sameness of person is identical to sameness of physical body or sameness living organism alone (Thomson, 1997, Ayers, 1936). The view expressed up to this point is a hybrid of sorts between the person-as-mental and person-as-physical positions. No mind exists apart from a body. Here 'body' can mean either a biological organism with a sufficiently developed and properly constructed brain or 'body' could be understood in a more general sense meaning some material object. Whether the mental emerges or supervenes it must emerge from or supervene onto something, and that something must be material. I do not think this view is, at first blush, controversial. The orthodoxy is that minds are tied to, connected with or stem from brains (at least any mind which exists beyond serious controversy). What makes the claim of this dissertation controversial is that the mind can be, in part, tied to, connected with or stem from non-biological bodies (like Otto's notebook) and can be external to the brain.

I will continue under the assumption that personal identity is a function of the mind with bodies, and not bodies alone. If we were persons based on body alone it seems a corpse would be a person. We, as persons, are our mental characteristic (which characteristic depends on exactly what theory of personal identity is at work), but these characteristics do not exist apart from the physical substance from which they emerged or supervened upon. If personal identity were sufficiently defined by body only the extension of cognition, as it has been discussed thus far, would not be possible. Consider a simplified view of memory. Let's say memory X is completely described and composed of a particular cluster of six neurons in a given physical, electrochemical state. These 6 neurons are necessary and sufficient for memory X to exist. If this were the case X could not extend, it could only be moved. All instances of X we encountered we would be sure that we are encountering the same token memory, not the same type of memory or an isomorphic copy of the memory.

If X were moved to a vat it would fail at least one of the criteria for cognitive extension; it would not be transparent in use. The amount of neural interpretation required to access the information therein would require far more effort than glancing at a notebook, asking the person standing next to you or checking one's watch. Further, we have not exactly 'extended' the mind, we have only complicated the means by which we access 6 particular neurons. Additionally, the cluster of neurons would fail to meet the criteria for endorsement once they are removed from our skull. This same information is encoded in the same neurons while in our head, yet we don't hesitate to call the memory cognitive in this case. The distinction is not because of the physicality of the memory but rather in what it is we are actually endorsing. When I recall the memory of my first day at graduate school I am not endorsing the exact neurons on which the memory is encoded (or reconstructed from). I am endorsing the memory itself by way of endorsing and trusting my internal, holistic, cognitive system. I did not choose to have this memory stored in the fashion and by the means that was. The memory was stored in the cognitive system by means which are, from my first-person point of view, unknown to me. This differs from encoding information in a notebook which is the intentional copying of (isomorphic) information from one endorsed system onto another endorsed medium by translating the physical, electrochemical state into language.

Maybe we could get past these objections if X were placed inside of another person's brain and incorporated in such a way the new brain could access the isomorphic information. Would this not be analogous to the divided and transplanted brain ala Parfit? Again, not quite, since the divided brain is the reordering of a single cognitive system by way of "stretching the wires," whereas the moving of the neuronal cluster changes the system in which the token

information exists. It is the same memory,[83] not an extended memory. Wouldn't we be inclined to say these 6 neurons are now part of this person's brain rather than an extension of the first person's mind?

For the mind to extend it needs to extend in such a way the isomorphic information is not literally the same physical instantiation (same token). We extend our mind to lessen cognitive burden or expand cognitive function. In doing so we are recognizing the limitations of the physical substrates of the mental tokens to be extended.

Otto and his notebook are a thought experiment designed to illustrate the extension of one kind of cognition: memory formation and retrieval. As such, I will here begin with Locke's (problematic) memory theory of personal identity (Locke, 1975) as a means of introducing the philosophy of Thomas Reid and Derek Parfit on which much of my argument will rest.

In short, Locke's theory of personal identity claims that a person is the same person as all the persons he remembers having been in the past. Though memory theory is not without its critics, this view is supported by, among other things, our intuitions rising from Locke's "body switching" story of the Prince and the cobbler[84] and the previously discussed (in the introduction) situation of a person not remembering a crime he is accused of committing.

---

[83] Perhaps it is not even a memory anymore. Sticking with the first day of grad school example, it is a memory so long as it remains within my cognitive system (extended or otherwise) but once moved into another's brain it is 1- no longer accessible to me and 2- at best a false memory as these events did not happen to the neuron recipient.

[84] The "body switching" trope is familiar enough that I do not believe it requires explanation here. IMDB.com lists nearly 100 movies and TV shows in which this trope is a major plot point; Wikipedia.org lists over 100 more. This does not include the use of this device in literature, of which Mark Twain's The Prince and the Pauper is a notable example. Locke's tale of the prince and the cobbler is the oldest example I have been able to find but there are, I have little doubt, stories in various mythologies involving this trope or at least something similar. (Wikipedia contributors, 2018, IMDB.com, TVtropes.com)

‘Identity’ refers to a relationship of being the same thing over time. But what is it that makes a thing the same thing over time? What are the criteria for this kind of sameness? Locke says that defining ‘identity’ is a difficult prospect as it is what he refers to as a “simple idea.” Just as it is difficult to define “blue’ without referring to colors or unhelpful physical explanations involving wavelengths of light, so too it is difficult to define ‘identity’ without referring to identical things. Difficulties of definitions aside, Locke does give three necessary conditions for identity:

1- When we see a thing (whatever it is) we know it is that thing and not some other thing

2- Two different things cannot be in the same place at the same time

3-One thing cannot have two beginnings nor two things one beginning. (Locke, 1975)

All three of these criteria have a place in a fuller conversation about personal identity but only the third of these is relevant to our discussion as it is the only one that addresses identity over time.

One major criticism of Locke is that his own explanation of personal identity violates this third criterion of identity. A person at time *T* is the sum of all the memories she has, one of which we’ll call *X*. If at a later time, *Q*, she does not remember *X* that person is not the same as the person she was at *T*. That person has ceased to exist. But if at *Q+1* she does remember *X* she exists again. And thus, the person has two beginnings which occurred at *T* and *Q+1*. The way out for Locke might have been to refer to dispositional memories. If she remembered *X* at *Q+1* she also remembered it at *Q* but was unable to recall the memory. But this is not a step Locke himself choose. In fact, on my reading, it does not appear this particular problem occurred to him. (Reid, 1855, Grice, 1941, Perry, 1975 and 2003)

I will examine Reid and Parfit's solution to this dilemma as well as throwing in my own two cents concerning cases specifically involving dementia patients who move in and out of lucidity, which I will appropriately, if not very cleverly, call the "lucidity argument". If Reid and Parfit are unable to convince the reader of the sameness of person of dementia patients in general I believe the lucidity argument at least gets the work done in matters regarding these cases.

In his opening lines Reid is in agreement with Locke in that:

> The conviction which every man has of his identity, as far back as his memory reaches, needs no aid of philosophy to strengthen it; and no philosophy can weaken it, without first producing some degree of insanity. (Reid, 1855, Perry, 1975).

The idea that memory is important in personal identity is obvious to Reid. In his Common Sense philosophy, he argues that one cannot even engage in basic reasoning unless one assumes one's memories are part of a continued existence.[85]

Reliance on memory alone has its limitations and Reid's solution to many of these problems is a relatively simple logical move. He adds a transitive property to Locke's memory theory which at the same time rewrites Locke's theory specifically and does much to preserve a role for memory theory in personal identity in general. The transitive nature of personal identity works thusly: Suppose I do not remember my tenth birthday, but I remember my eleventh. By Locke's reasoning I am not the same person now as he who had a tenth birthday, but I am the same person as the person who had the eleventh birthday. However, say, on my eleventh birthday I did remember my tenth birthday. This means that I, today, remember being a person

[85] If I want to solve the equation '(2 + 4) + 6=' I have to first solve the parenthetical. Once having done that I can add the remaining integer. However, if I do not assume I was the person who solved the parenthetical, I cannot know if the answer I now hold can be trusted, so I have to go back and check. But if I do not believe I am the same person who went back to check, I cannot assume the answer I now hold is correct, so I have to go back and check, etc. ad infinitum.

who remembers the tenth birthday and thus personal identity is continuous between me today and me on my tenth birthday.

Reid's claim that personal identity is transitive is not necessarily an obvious one. A person is changing, physically and mentally over time. The fact that a human life exists linearly does not guarantee transitivity. "Mother of" for example is not a transitive relationship though a series of mothers can be a linear chain. In my family tree, Nancy is the mother of Theresa and Theresa is the mother of Katie and Katie is the mother of Allyson[86] but it does not follow that Nancy is the mother of Katie or Nancy is the mother of Allyson or Theresa the mother of Allyson. (However, there is a transitive relation to be found in the term "matrilineal ancestor of".) On what basis does Reid argue that personal identity is best defined as a transitive relationship rather than any other linear one?

Where Reid and Locke differ is that Reid does not believe that continuity of personal identity rests in memory (or any other psychological phenomena) *alone*. Whereas, Locke is ready to separate the human animal from the person, Reid sees this as an impossibility as the two are inextricably linked. The human animal, for Reid, plays an important role in the continuation of the person. Without saying so outright, Reid argues that without the animal the person cannot exist. The same continued life, Locke says, accounts for the sameness of animal over time, and Reid agrees. Animal identity remains the same over time, i.e. – I am not born one life form and die a completely separate lifeform. My life is the same life so long as I am alive. At any moment which the human animal is possessed of personhood that personhood exists in the same transitive relationship to all other moments of personhood through their connection with the same animal. My tenth and eleventh birthday are both stages in the same animal life as I am today. Though

[86] My grandmother, mother, sister and niece respectively.

what and when is remembered changes from time to time all memories are nonetheless the product of the same animal existence and are tied to whatever person happens to exist at a given moment.

Parfit discusses this issue in Reasons and Persons. Here Parfit illustrates an important distinction which escaped both Locke and Reid; the distinction between psychological connectedness and *mere* psychological continuity. It is the 'mere' in this conjunct which Parfit will argue holds an important key to personal identity.

Since identity over time is transitive (else there would be a break in continuity and thus a violation of Locke's third tenant of identity) so to must be any criteria for personal identity also be transitive. Psychological connectedness occurs when a person at time *T* has a memory of events at time *T-1*, such as I have memories of things I did yesterday. I may also have memories at *T* of *T-2* (the day before yesterday). However, people generally have only a few strong memories of *T-20* years, (though there may have been strong memories of *T-20* years at *T-20+1*) that is to say at *T* I am not psychologically connected to a person at *T-20* years. Though we can run together a backwards string of psychological connections (*T* to *T-1* to *T-2 to T-3* ... to *T-20* years) this relation is not transitive if we are considering just connectedness. The connectedness does not survive this continued backward narrative. The events of my past which I do not remember did happen but are not a part of my life today. (At least not consciously. There are certainly unconscious effects.) This chain does not make it the case I have at *T* memories at *T-20* years only that existed at *T-20* years. Meaning, this same narrative is transitive in that it shows an unbroken chain of psychological states and thus demonstrates psychological continuity despite my lack of psychological connection to the person at *T-20* years.

Though I do not myself hold the belief that personal identity is based on a millisecond by millisecond chain of continuous psychological states such a view is not impossible. Memories and other mental states chained together on an instant to instant basic form what Parfit calls 'psychological continuity. This is a weak claim about mental lives which tells us next to nothing about one's personal identity. I, for example, would not explain who I am as person by recounting the technical minutia that binds my mental life into a coherent whole. I would, on the other hand, relate to you the big events and important facts which are, to me, part of who I believe myself to be. This narrative is what Parfit calls psychological connectedness. There is a psychological continuity between me today and me on November 24, 1987 (a date chosen at random on which I was alive but of which I have no readily available memory, or maybe I do and I just do not associate those events with this date) but I am not necessarily psychologically connected to mental events of that date. Today, however, I am psychologically connected to the mental events that occurred on September 19, 2015, the day my son, Henry, was born. Though the events of both these days are part of my continued existence, my relationship to them is very different. I will argue that mere psychological continuity becomes more akin to psychological connectedness when the extended mind is considered, and advance directives enter the picture.

One reason psychological connectedness becomes mere psychological continuity over time is that fact that the human brain is pretty lousy at both long- and short-term information storage. The amount of information around in the world vastly outstrips the central nervous system's processing ability. Further, information which is processed is judiciously sorted and selectively stored for future use. We remember very little about what we experience and what we do remember is subject to revision and reinvention (Dennett, 1991, Gangopadhyay et al, 2010). Over time the fidelity and accuracy of our memories decline. If information proves to be

unimportant, or goes unused, we tend to not spend our limited cognitive resources keeping it available. Extending our minds helps get us around these limitations. Once Otto commits information into his notebook that information is no longer subject to the same degradation over time as it would be were it in his wetware.

Back in chapter 2 I mentioned how few phone numbers I have committed to (biological) memory. I do not, at present, have my sister's (Katie or 'Beans') phone number in my biological memory, instead I have it stored in my mobile phone. Before I had such a phone I did remember Beans' number by heart, but all I can tell you now is that it has a 248 area code and I think there are a couple of 6s in it.

At some point in the past I learned my sister's phone number, but I do not remember having done so. This information is a semantic memory in that there is an absence of an episodic memory explaining where the information came from. While I remembered her phone number using only my wetware, we could make the case that I was psychologically connected to that particular episode of my life despite not currently having any direct memory of the incident because there existed a result of that episode (having the phone number readily available) and therefore particular consequences and outcomes as a result. If I forgot her phone number those results would disappear and be replaced with different consequences and outcomes (i.e. - things about my life would work differently under certain circumstance).

We could draw a backward narrative chain from myself currently (a) to myself in the past when I remembered the phone number (b) to me when I learned it (c) thusly (where 'R' means 'is in relation to'):

$(Rab \rightarrow Rbc) \rightarrow Rac$

This narrative describes the psychological continuity of these three stages of my mental life bound together by my continued animal life. It was by this that Locke argued I am the same person over this course of time. But as Parfit would later point out, I do not, in fact, remember (the phone number) now at (a) which I learned at (c) as the above symbolization might suggest. Memory and animal identity do not always survive transitivity together. The fact that this information was remembered by (b) and the (a) and (c) are in relation to (b) does not mean (a) and (c) are possessed of the same content. There is a chain of mental continuity but no connectedness. If we consider this fact, the relationship is properly described as (where 'M" means 'remembers'):

(Mab & Mbc) & (~Mac v Mac)

(Mac) is possible but we have already stipulated it is not the case, therefore:

(Mab & Mbc) & ~Mac

My lack of memory means this is a non-transitive relation much like: humans eat cows, cows eat grass, but humans do not eat grass. In this description of my relationship with myself over time the memory of the phone number has disappeared. Since it is not a memory I now have, the fact that my particular biological animal remembered it in past is causally impotent. It makes no sense to say that because I remembered the phone number in the past that I am today mentally identical to that past person if I do not have actual access to past mental content. If I did remember the phone number today, we could eliminate the need for (b) entirely:

(Mac)

Since (b) it is not causally necessary[87] for me to remember the number at both (a) and (c).

[87] All that is necessary is that some stage(s) or other filled in the intervening time.

What if I save my sister's phone number into an address book or my mobile phone? The mental content of my sister's phone number has been extended into an external (relative to my wetware) medium and therefore it is not subject to alteration or being forgotten in the same way my biological memory may be. If my phone meets the criteria for the extended mind, I will always remember this information.[88] Once I extend my mind what counts as a transitive relationship changes. The relationship between my past selves and my current self becomes transitive. I remember the phone number, not just remembering being someone who remembers the number.

The Lucidity Argument

Like the above transitivity argument this "lucidity argument" is concerned with identity as it relates to Locke's third principle of identity (but, again, elements of the first and second are at play as well). The progressive nature and often intermittent presentation of dementia means the cognitive faculties of the patient are hard to pin down and this can make a determination of identity difficult. Loss of cognitive function such as memory, facial recognition and speech happen over (perhaps a great deal of) time. A patient showing no competence or understanding of her current situation at one moment may be reasonably lucid and somewhat "back to her old self" the next.

Let's consider a patient (H) over time. For the sake of argument, we'll assume that until the onset of noticeable and not easily dismissed symptoms (H) had a personal identity which remained unbroken throughout her adult life. As her symptoms progress she sustains longer and longer periods during which she is unrecognizable to those who knew her as the person she was. That is to say, much of the time the patient seems to have become someone else.

[88] So long as I have the right kind of relationship to my phone.

When lucid, (H) is identical with the person she was when she was healthy. While exhibiting symptoms of her condition, she appears to be someone else, call this person (D).[89] Given the presentation of what appears to be two distinct persons, we can conclude that continued life (L) which was once the locus of (H) alone has instead become the locus for (D) was well. At some point during (L)'s history (D) first presented itself. The same is true of (H). (H) began at the earliest instance of her psychological continuity. Locke's first and second principles of identity are that when we see one thing we know it is one thing and not something else and that two things cannot be in the same place at the same time. So here, when we see (D) we know it is not (H) and vice versa.[90]

When (D) is present, what happened to (H)? If (D) is "always on" we may be able to say (H) has simply ceased to exist, but this is not the case here. (L) has moments, perhaps only fleeting, but moments nonetheless, where (H) is present. If this is the case, (H) must have existed all along else we run afoul of Locke's third principle. If we do not say (H) has always existed, we can only account for her "reappearance" by either claiming:

1- (H) ceased existing and later began to exist again.

or

2- The reappearance of (H) is not an instance of the same person but instead a new person who is qualitatively the same as (H).

I wish to dismiss option 2 out of hand. It is a far simpler explanation that (H) continued to exist than the same pile of brain matter resulted in two different instances of a qualitatively same

[89] It is likely that (D) is not a consistent, single, unified person over time but rather a successive series of short-lived persons (D_1, D_2, D_3 etc.). But for the sake of simplicity we'll assume (D) is the same. Parallel reasoning applies in either case.

[90] This is also assuming (L) cannot be two persons at the same time (else Locke's second criteria for identity has been violated).

person. Option 1, I believe, is just as unlikely but requires a little more explanation as to why this is so

While (H) was living her life she had memories, one of which was of the day her beloved dog, Spot, was adopted from the animal shelter. What made it the case that (H) had this happy memory was the fact that it is based on events that actually happened to her. She remembers it because she was there to form memories of the event.

After (D) begins making appearances we are faced with a problem if we believe above option 1. When (H) reappears what are we to say of the memories of Spot's adoption? The person, (H), who remembered that day no longer exists; (D) killed her (so to speak), yet there appears before us a person claiming to fondly remember Spot. If (H) is gone the person claiming to be (H) only came into being very recently. As such she was not present when Spot came home. The "memory" of Spot is not the result of being present to have experienced the events but is instead some kind of a false memory. Since the person claiming to be (H) only came into existence a few moments ago she cannot truthfully claim to be the person who remembers being the person who remembers Spot's arrival. No kind of transitive relation among memories exists at all.

In the end neither 1 nor 2 is the case, so it must be that (H) has indeed existed all along. How can this be? An alternative explanation is that (H) has always existed. After (D) begins presenting itself we should not think of (H) as disappearing and reappearing, instead (H) recedes and resurfaces. She has always been there. (H)'s personhood has been a constant, just hard to get in touch with at some moments more so than others.

The lucidity argument is not as far-fetched as some might think. In fact, we already engage in this kind of assumption when dealing with all human persons as they go through a

cycle of competency and incompetency on a (nearly) daily basis. We call it sleep.[91] While sleeping we are temporarily incompetent but we do not think the sleeper qua person has ceased to exist. This easily contrasts with the demented patient sliding into, and out of competency. However, there is a difference between the competent sleeper and the occasionally competent patient. While sleeping we are the same person but are unconscious and unable to act or express desire. The incompetent patient will be, at times, unconscious (either sleeping or comatose) but at other times conscious and making demands/requests and generally behaving in a way as a person of diminished capacity might.

The short story of the argument here is that the patient remains the same person throughout the progression of the dementia so long as (H) can/does reemerge in moments of lucidity. But how do we prove the negative? When do we say (H) cannot/will not reappear? It is not logically possible to make such a claim so long as the patient's brain retains any kind of meaningful cognitive function. Thankfully this need not concern us in the case at hand. A failure of (H) to resurface is not evidence of her non-existence. The continued existence of (D) reaffirms (H)'s existence.

I argue that (H) never fully disappears, even if she has forever receded into the background. Even when (D) is at her fullest and most demented state, (H) is still playing a role in (D)'s mental life. (H) is a necessary condition for (D)'s very existence. Both (H) and (D) are the result of the same cognitive hardware; they share the same brain.

---

[91] The following analogy not only applies to sleep but also any other loss of consciousness or states subconsciousness (such as sleepwalking, fugue states, intoxications, hard hits to the head, etc.). Different moral standards may be applied to the behavior of a subconscious person, however. Getting hit on the head in an accident might mean you are not liable for your actions, similar to what we would see in cases of legal temporary insanity, while intentionally getting black out drunk is not an adequate defense absolving one of the moral and legal ramifications of her actions while in such a state.

Imagine (D) is refusing to sit still so that her nurse can insert an IV for antibiotics, because (D) is afraid it will hurt. (D)'s right. Needle sticks are unpleasant, but this disposition is not innate or inborn. (D) had to learn this fact about needles from somewhere, and (D) learned this because (H) learned this. (D) is reacting uncooperatively to a pending needle stick whereas (H) would be more likely bracing herself for it. These are reactions to the same semantic memory; the memory which was encoded when it was learned needles are correlated with pain, a memory which is psychologically continuous with (H) today. This memory is, transitively or directly, (H)'s memory. (D) is using this memory, or rather (H) is, through (D), but applying the memories differently.

To push this thought experiment a little farther we might say that (H) and (D) are not different persons at all. Rather than different persons, we can think of them as different manifested outcomes based on the same input (memories) resulting from the application of differing rules of interpretation.

By analogy every reader versed in base 10 arithmetic would determine the outcome of the addition problem 15+18 to be 33, and that answer would be the same so long as 1- the numbers and meaning of the operators remains the same, and 2- we continue to apply the rules of base 10 arithmetic. However, if we consider 15+18 and apply the rules of a base 9 system the answer is 36.

Both 33 and 36 are valid answers so long as you are evaluating those answers against the number system by which they were derived. Here 33 and 36 are the same number, though represented in different characters which mean different things depending on the base counting system you are dealing with. If I am adding in base 32 and you are evaluating in base 10 you

will judge my result of 12 to be in error, but you are the one who is in error from my point of view.

Getting back to (H) and (D), they are both pulling from the same cognitive resources (memories, dispositions etc.). Any actual memories that (D) might be drawing on are numerically identical to those memories of experiences encoded by (H) at some time in the past. Any false memories or delusion (D) might drum up are drawn from the (H)'s resources as well. If (D) is having delusions of black helicopters surveilling her every move the presence of helicopters in this delusion is the result of (H)'s previous knowledge of helicopters.[92] (D)'s delusion of being watched may be false, but it is not 'wrong'. It is false insofar as we evaluate her cognitive process from the stand point of a competent person but given the input (memories of helicopters) and the rules of operations (the altered brain states resultant from a damaged brain) the result is the 'right' (i.e.- expected) result in that (D) believes it. The most we might be able to say is that (H) and (D) are different kinds of persons. A Wall Street executive who decides to give up his seven-figure salary and become a beet farmer in rural Idaho is at each time a different kind of person from who he was at the other time, yet he remains numerically the same person. Even if (D) exhibits only the worst possible traits a person can show, it does not mean (D) is a metaphysically different person, just a super unpleasant manifestation of the same cognitive content.

Determining Sameness of Persons

The surest way of evaluating psychological continuity requires the reporting of the subject with respect to whom the continuity is in question. Knowing if someone remembers

[92] It could be the case that (L) while (H) never experienced helicopters but (L) while (D) had. This is merely the same problem in reverse. Anytime (H) recalled helicopters she would be drawing on the same memory as (D) which was formed by (D).

what happened two minutes, two months or two years ago, with as much certainty as possible, requires asking that person for a report. But we do not, under normal circumstances require this level of certainty nor do we routinely call into question our own assessment of the psychological continuity of others. If you and I met 2 minutes ago, I know that you remember having met me so long as there is no reason to think that you have memory issues. I don't need to ask for a report in this case, but I am nonetheless assuming your sameness of person, though not explicitly evaluating your psychological continuity. That being said, the questions of reasonable doubts as to the numerical identity of such patient. As such, the need for first person reporting will be greater.

The extended mind thesis does not eliminate this need for first hand reporting of what one remembers, it does create new avenues for getting at this otherwise first-hand information. Insofar as Otto's notebook is really the contents of his mind, which are observable to the outside world, his entry stating, "on October 3, 2009 I went to Disney World" is just as informative to us as him orally telling us. So too are the contents of an advance directive – the observable contents of a patient's mind; in this case her wishes for her future medical treatment. Whether read off paper or contemporaneously reported, the information is the same, yet there are two differences which need to be consider. First, whereas the note about having been to Disney World is a report about a past event which either is or is not true, the wishes expressed in an advance directive are concerned with future events. Second, is the issue of consequence. Stating in writing that he went to Disney World on a given date is not likely going to be of any (serious) consequences for Otto.[93] An advance directive is, by its nature, a document of consequence. What's more, these wishes persist throughout time regardless of the functional state of the patient's biological brain

[93] Maybe it might help him establish an alibi.

and retain causal potency. The statement "on October 3, 2009 I went to Disney World" just a reporting of a fact, it is causally inert in that this past fact of Otto's life will not likely affect what will happen to him in the future.

If an advance directive is a kind of mental phenomena we need to consider what particular kind of a mental phenomenon it is.

In the following section I will argue that advance directives are performative acts. This position is not new. Buchanan and Brock make a similar argument:

> The issuing of the advance directive is a *performance,* [emphasis original] and act of will, in just the way in which drawing up an ordinary will for the disposition of one's property is a performance, not a mere expression of preference. Similarly, when a competent patient deliberately and unambiguously refuses life-sustaining treatment under conditions of informed and voluntary consent, his refusal is not merely evidence that he prefers not to have the treatment; it is, under these conditions, an act of will. (Buchanan and Brock, 1990)

What I wish to add is that cognitive extension can be applied to performative actions because they conform to the previously discussed criteria that candidates for extended cognition not be only semantic, but specifically isomorphic in nature. If this is the case, the presence of an advance directive, qua performative act, creates psychological connectedness (not mere continuity) between the patient in the past and the patient in the present. Though we are not necessarily bound for our entire lives by previously executed performatives. I am not bound by my high school promise to Shannon Crawford to marry her when we turn 40 if we are not each already married to someone else. Not only am I married and over 40 years old, but the promise was made when we were teenagers who a few years the promise was made Shannon and I lost touch. As circumstance change, so too does the appropriateness of carrying out promises. Advance directives are a performative of a special type not susceptible to the possible failures and misfire or unhappiness discussed by Austin (Austin, 1975), as we shall soon see.

Advance Directives, Ethical Kinds and Performative Action

Whether it is Locke, Reid or Parfit, the connectedness (or continuity) of a person has all been based on connectedness (or continuity) of *internal* mental states; Reid having gone so far as to say that it was "insanity" to think otherwise. The possibility of an extended mind raises questions as to how we should view this relationship in terms of what is continuous or connected to what. If an advance directive is indeed part of a person's extended mind does it apply even if she does not remember drafting it or the contents of what was drafted?

There is a quote from Buchanan and Brock's Deciding for Others which is all but tailor made for discussions of the extended mind thesis and advance directives:

> … if the concept of a person implicated in the psychological continuity view is a purely metaphysical, nonmoral concept, then it is hard to understand why certain marks on paper made by a person in *this* (metaphysical) sense should be thought to create obligations or confer authority, since obligations can be created and authority can be conferred only by persons in the moral sense. (Buchanan and Brock, 1990)

Buchanan and Brock argue that an advance directive is properly seen as a performative act (Buchanan and Brock, 1990). Advance directives are future regarding documents. Buchanan and Brock are concerned with how advance directives gain their moral force at the time which they are needed. The causal potency and cognitive relevance are only in force when implementation or interpretation is needed. I will go farther and say that, like any other dispositional mental state, advance directives are part of a person's cognitive machinery after they are drafted and until they are no longer applicable (after the person has died or changed/rescinded the directive). My memories of how to tie my shoes are not only part of my cognitive machinery when I tie my shoes. This memory is always present, just not necessarily occurrent. Insofar as advance directives are mental states, this must be true of them as well.

I also argued in chapter 2 that the extended parts of extended cognition were only those parts which are isomorphic in nature. So, are performatives semantic[94] or isomorphic? The answer is that they are both depending on which parts of the extended cognitive machinery you are looking at. The extended parts of cognition which contain only information contain isomorphic information. The semantics of the performative are found in the biological brains of those interpreting the information. The argument is reasoned in a parallel manner to the previous arguments regarding original or non-derived content in previous chapters.

The first step of this argument is to look briefly at John Austin's work regarding linguistic acts. Performatives are a special kind of statement which in their utterance preform an action. "I do" is an example of consenting to the vows in a wedding. "I christen this ship the …" is a linguistic act that thereby names the ship. "I promise to mow the lawn Monday" is the act by which one promises to mow the lawn on Monday. The performative act does not occur in isolation and performative, when fully described, are not just the words uttered. The mental states and decisions which lead to the utterance are indispensable for a performative act to take place and are thus tied to the output which constitutes the act itself. For the act to be meaningful, binding or what have you, there must be semantic content in the act, that is to say the act must mean something. The words themselves are just the end result. I can run around saying to everyone I meet "I do" but this does not mean I have married all these people. The situation is wrong. My many "betrothed' are not in the right frame of mind to understand and assent to what I am doing. The words "I do" in the right context, with the right players, are the end result of a process full of meaning and matter to those participating, as Austin well understood. Likewise, I could utter the Greek phrase ΘΗΜΑΜΑΙ ΠΕΡΗΣΙ (pronounced thim-

[94] That is to say, semantic and not isomorphic.

amĕ per-si) but it would not be a true performative. I do not speak nor read Greek. This example was provided by my friend Michael Zois. I do not know what it means. Just in case it's something dirty (not an unrealistic possibility), I will utter a performative I do understand: "I hereby disclaim responsibility".

Though marriage is a complicated contract requiring semantic understanding of the relationship being entered into, it does not follow that the act of consenting to marriage qua act is itself a semantically filled act. The words "I do" are nothing more than an acknowledgement or acceptance of the semantics which preceded. The semantics still exist (in the marriage contract and in the idea or understanding of the ones getting married of what a marriage is) and need not be found in the output of the linguistic act itself. The words are just the means by which we perform the action.[95] Imagine a person who is both deaf and mute who is getting married. He cannot hear the vows nor speak the reply, yet he is just as able to get married if the vows and response are given in a written form or by giving a confirming nod. If there is no semantic content in a verbal reply itself, should we not hold a written reply to a different standard? The written 'I do' is just as binding and just as much a performative act. This is because the information is isomorphic between cases, assuming relative sameness in the performer's grasp of the semantics of the situation.

The extended mind thesis changes our view of what an advance directive is. The contents of Otto's notebook are his thoughts and memories stored outside his wetware. The

[95] Of course, the words "I do" can be uttered under duress, coercion or incompetence. If we know this will be the case before the performative occurs, we know in advance that we will not accept the linguistic act as legitimate. If this fact is only known after words are uttered, we then change our previous disposition of acceptance of that of denial. See Austin's extensive discussion of "misfires," "infelicities", and so forth.

notebook is not a second best, copy of mental activity; it is his mental activity. The same would hold for any performative, linguistic act of which there is an external observable record of a mental state, such as speaking out loud. A directive is a performative act by which the drafter makes his wishes known to future readers. The directive is the drafter saying to the reader "I hereby request that should circumstance X obtain it is my wish that Y". The advance directive is an observable part of the mind of the directive's drafter.

If the person who drafted the directive is the same as the patient to whom it now applies, the advance directive is the performative linguistic act of the patient as much as it was of the drafter. But does it necessarily apply? Should we always enforce the instructions of these documents? How we answer these questions in light of the extended mind will depend on two factors. The first is what kind of metal state or frame of mind is the patient currently in. Whether the patient is pleasantly or otherwise demented will change our obligations to the wishes of the patient. And secondly, an appropriately placed proxy may adhere to or override the directive with as much authority as the patient herself has. The first of these factors will be discussed under the following headings of this section, while the second will be covered in the following chapter dealing with interpretation problems of advance directives.

In addition, there is reason for thinking we should take these extended desires more seriously than other performative actions. Advance directives are unique in that they are not vulnerable to at least one of the misfires that Austin says can make a performative "unhappy." Although we all know people have made promises with no intention of keeping them we still rightfully feel wronged when such a promise is broken. We have been lied to. As Austin says "[a]ccuracy and morality are on the side of the plain saying that our word is our bond". The 'unhappiness' of an advance directive can only be brought about by the failure of its being

carried out. An advance directive can be made unhappy by a failure of the instructions to be carried out at the appropriate time, but the drafter is incapable of making her own directive unhappy.

When I say "I promise" with no intention of keeping this promise it is not the case I did not promise but rather I promised in bad faith. If I draft an advance directive I cannot do so in bad faith. Insofar as an advance directive is a linguistic act, it is an act of the utmost importance. The stakes are very high; literally life and death in some cases. If I were to draft an advance directive which I did not really wish to be carried out should the correct circumstance obtain, this would be a sign that I am not competent to draft such a directive in the first place. If I am competent and draft an advance directive the nature of the directive itself assures my honest and complete endorsement (criteria 4 of extended cognition) of what it contains.
Though some interpretation problems may occur when the time comes to execute the directive (again, a problem for the next chapter) there can be no doubt that the wishes on the paper are in fact the wishes of the drafter.

The nature of an advance directive being a future regarding choice creates a problem. It is an act of will consenting, or not, to particular medical procedures should the appropriate circumstances arise. In the intervening time between the drafting and the need for implantation circumstance could have changed. In chapter 1 defined four different kinds of complications which could arise in this intermediary time, which I collectively called KPADs (Knowledge Problems for Advance Directives). The first KPAD discussed raised questions as to whether, in some cases, the patient now is the same person as the person who drafted the directive. It is inappropriate to apply one person's advance directive to a different person. If advance directives are to be legitimately carried out there needs to be an argument that either 1) the present patient

is the same person as the past drafter or 2) the wishes of the present patient are rightfully subordinate to the past drafter. I will discuss the implications of the extended mind thesis as it relates to the change of person which worries Dresser in three forms: no person is present, a pleasantly demented person is present, and a severely demented person is present.

Change of Person- No Person at All

Does death, or loss of personhood, mean nothing of the person remains? At the very least the body of that person persists, but as some say colloquially this is "just a shell" and not the same thing as the person who once was. What I mean when I ask "does anything of the person remain" refers to anything remaining which made a person a person (in the metaphysical or ethical sense as discussed by Rovane, 1998 which will be discussed shortly)? Insofar as a person is some subset of her mental characteristics and insofar as an advance directive is a persistent mental phenomenon, it may be the case that something of the mind of the person remains despite the physical condition the patient finds itself in.[96]

There is much loaded into this answer to the question of "what remains" as well as a number of unsupported assumptions. It is not just any mental state that we think divides the line between human person and non-human person. Near the end of her biological life Terri Schiavo no doubt had some very low level mental states (likely originating in the reptilian brain or other primitive neural structure) but she nonetheless fell far short of the mark of 'person' as these states were well below the threshold of conscious awareness. Stimulus response is a cognitive function, but any living organism is capable of them. The fact that Schiavo slowly turned her

[96] There already exists a dispositional mental state of this type (a persistent mental phenomena) we treat just as we would the contemporary utterances of the one to whom the mental state belongs: a last will and testament. Though our treatment and execution of these documents already matches the following prescriptions for advance directives, last wills would also benefit from the supports of the arguments made in this dissertation.

head to follow the path of a moving object is itself no more a sign of personhood than a maple tree turns its leaves to better collect sunlight. What is needed is an explanation of what sorts of mental states make a person a person. Further we need to determine whether or not these mental states can survive into the advance directive document.

Carol Rovane argues that 'person' is a phenomenon rightfully the subject of at least three areas of inquiry: science, metaphysics and ethics. The physical remains of the person are what is left of the person of interest to the biologist, pathologist archeologist or coroner, but is there anything left of interest to the metaphysicist or ethicist? In light of the extended mind thesis I will begin with a tentative yes.

Rovane gives an ethical criterion of personhood stating:

> …something qualifies as a member of the kind 'person' *just in case* (call this 'P') it has the ability to engage in agency-regarding relations. These are relations in which one agent attempts to influence another and yet aims not to hinder the other's agency (call this 'R'). (Rovane, 1998)

[Parentheticals and emphasis are mine, which are to be discussed in the following section titled 'Rovane's Ambiguous Conditional".]

Rovane acknowledges the moral importance of the fact that some human behavior is aimed at modifying the behavior of those around us. This is not as insidious as it might sound. Even something as simple as asking for a favor would qualify as such an attempt. If I ask you to "please pass the salt" I am attempting to modify your behavior, yet I have done so in a way that does not 1) fail to treat you as a person and 2) does not hinder your ability to make choices (i.e.- hinder your agency). In asking you to pass the salt I have not coerced or manipulated you into doing so. I am attempting to influence your behavior but in a way that treats you like a person and respects your autonomy while still attempting to see my desire for proper seasoning fulfilled. Whether or not you pass me the salt is a decision you make freely. I would be grateful if you

passed the salt and annoyed if you did not, but nothing in this exchange diminished the capacity of either party to act as free agents. This is, in Rovane's sense, an ethical exchange between things of the kind 'person.' If you do not pass me the salt you have, at worst, breached a rule of etiquette. Assuming you didn't refuse to pass the salt due to a dismissal of my, or anyone else's personhood, the exchange remains an ethical one regardless of how it unfolds.

If on the other hand an 18th century plantation owner tells a slave to "pass the salt" he has not engaged in an agency-regarding relation since he has a prejudice against the slave manifested by his denial of the slave's status as person. The slave does not, in any pragmatic sense, have a choice whether or not to pass the salt. The slave is not a person, in the slave owner's eyes, as he is expected to carry out the task assigned regardless of his own wishes.

An advance directive is an attempt at agency regarding relations. It is an expression of wishes as to how the drafter wishes others be to treated or be regarded by others. The advance directive is an attempt to influence the behavior of others ("I hereby request that you do this for me, as I am unable to do this for myself"). If a person is a kind of thing which engages in agency-regarding relations and an advance directive is an engagement in agency-regarding relations, it stands to reason that the presence of such a directive indicates the presence of a person.

There is an obvious objection to this syllogism. Does it really follow that the presence of an advance directive implies the current presence of its drafter or just a drafter in the past? The ashes of a campfire do not imply the current existence of the person who started the fire, only that she existed at the time the fire was started. Nor does the continued existence of the pyramids at Giza imply the builders are still around to bask in their accomplishments. Are we to say the person exists beyond her biological death because her advance directive was written on durable

paper? There is a substantial literature on the issue of surviving interests and harm to the dead (Nagel, 1979, Partridge, 1981, Feinberg, 1984, Pitcher, 1984, Marquis, 1985, Callahan, 1987, Dworkin, 1994 to name but a few). It is not necessary to give a complete overview of this material. A summary in the form of the following thought experiment will suffice:

Imagine a theory is put forward that the Harry Potter book series was not written by J.K. Rowling but rather generated by a computer. Such a claim (should it gain traction within the population) would harm Rowling's reputation and therefore harm Rowling. It is assumed that because Rowling was harmed Rowling must exist. I cannot, after all, harm unicorns, Batman or the married bachelor.

If we apply this same reasoning to the historical case of William Shakespeare can we get the same result? There is a body of scholarly work which claims the corpus of The Bard was actually written by Christopher Marlowe.[97] This (if it gains traction within the population) harms Shakespeare's reputation and thus harms Shakespeare. Since Shakespeare is harmed, he must exist.

Rowling, as of the writing of this sentence, is alive and well *living* in England. Shakespeare, as of this writing, is still very much *dead and buried*, in England. Can we really say both authors are harmed, or at least harmed in the same way, by the slander calling into question the authorship of their respective works? In its face this seems preposterous. How can something be harmed if that something does not exist? Shakespeare has been dead for more than 400 years. He cannot experience harm or suffering, nor praise or pleasure.

[97] Various forms of this conspiracy theory are called the "Marlovian theories of Shakespeare authorship". If one were so inclined wikipedia.org would be as good a place as any to start your research.

This is a very fair point. But just as one can have interests of which one is not aware, so too it can be possible to be harmed and not know it. Taking my cue from Nagel (1979), consider the state of having been betrayed. To be betrayed is a harm. We don't like it when we have found out we have been betrayed. But it is not the finding out about the betrayal which is harmful. It is bad to have been betrayed even if we never find out about it. Arguments can, and have been, made that being dead does not change the possibility of being harmed (Nagel, 1979, Partridge, 1981, Feinberg, 1984, Dworkin, 1994) as you are just unaware of the situation for different reasons.

The possibility of the extended mind creates another possible argument for harms despite a lack of awareness of them. All persons integrated into a social unit leave behind some trace of themselves when they die or otherwise cease to be a person. There are relationships, written works and, if nothing else, a corpse.[98] But an advance directive is a different kind of thing to leave behind. It is an act of will. An advance detective is a persistent mental state aimed at controlling one's own future. This controlling comes in the form of an agency-regarding relation. Rovane's ontological claim about the kind 'person' as an ethical kind coupled with the claims of extended mind theorists make a strong case that the presence of an advance directive indicates the presence of a person, albeit in a form we are not accustomed to recognizing. Persons are an ethical kind and an advance directive is an externalized manifestation of the kind of thing that make a person an ethical kind, an attempt to engage in an agent-regarding relation.

Does an advance directive apply if the person who drafted it no longer exists despite the continued life of her human organism? The conflict we face is this: we have in our possession

[98] Though, we could concoct some rather gruesome scenarios involving high explosives, a vat of acid or scavenging animals where even this is not the case. However, the matter which made up your body would still be around in one form or another.

an external, observable, persistent mental state (the advance directive), and mental states have to be the mental states of some person or other, yet the current orthodox definitions of person would have us believe that the person to whom the directive belongs no longer exists.

If the advance directive is viewed as something less than actual mental phenomena it has no morally necessary role, as regards the drafter, in determining the future treatment of the patient as the person for which the directive was intended no longer exists. In the absence of recognition of the directive as mental phenomena it is just words on paper for whom the subject no longer exists. Applying the directive to anyone, at this point, is akin to forcing the past will of a now extinct person onto an existing person.

This concern is not warranted in this case for two reasons. First, we are not discussing here whether or not to apply this directive to a different person (this will be done in the next two sections) but rather to no person at all. In this case then patient in question is so far gone mentally the title 'person' does not reasonably apply.

And secondly, if an advance directive is a mental state—an ethical mental state for that matter— (as a synthesis of the extended mind thesis and Rovane's argument of 'person' as an ethical kind) and persons are an ethical kind the result is that something of the person remains in the form of the advance directive after the end of the person's biological life. The presence of a competently drafted advance directive means that it is not true nothing of the person remains. What remains, and maybe all that remains, is quite relevant to the state the person now finds herself in.

As stated earlier (with the promise of an argument forthcoming in chapter 4 in the section titled *Isomorphism in Language and its Limits*) the advance directive on paper is an isomorphic mental event. The semantics decision to draft a directive of this form occurred before it was

written. The fact that the person no longer exists does not change the fact that a person, through the use of isomorphic thought, drafted this syntactical record of those thoughts.

If the advance directive is the only remaining mental state, there is no current metal state for it to contradict or be in conflict with. The directive is the only mental state to consider and thus it should be applied. To do otherwise is to make the performative of the patient unhappy. The only mental state relevant to the personhood of the patient which remains is the advance directive. To not carry out the directive is to fail to respect those parts of the person which remain.

Rovane's Ambiguous Conditional

How we parse out Rovane's definition of 'person' as an ethical kind matters not only for what qualifies as a person but also can affect how we view advance directives in light of the extended mind thesis. The concern is whether she means her definition as a one-way conditional or a biconditional. In this section I will briefly explain a plausible interpretation of each kind of conditional and conclude, for the sake of argument if nothing else, that the one-way conditional was intended.

The phrase 'just in case' is translated in logic as 'if and only if', which I find to be problematic here depending on how strong a claim Rovane wishes to be making.

Being a biconditional ($P \equiv R$) the stronger reading claim is:

1. Only persons can recognize and the attempt to influence agency.
2. Recognizing and attempting to influence agency are required for personhood.

This is a more limited use of the word 'person' than I usually endorse; it is far too narrow a definition. It leaves out many, if not all, non-human persons. Non-human animals are capable of

behavior which are appropriately and most easily described as 'ethical'. [99] Additionally, a person with autism is certainly a person based on any previously discussed criteria for personhood yet may fail Rovane's test. He may engage in behavior aimed at influencing the agency of others while nonetheless being unable to grasp or recognize that agency.

The implication of a biconditional, such as ($P \equiv R$), means that both P and R are necessary and sufficient for personhood, but given the two counter examples above this is clearly not the case. Rovane might mean something slightly less strong, however.

There is a reasonable (though technically incorrect) interpretation of 'just in case' (Temple, 2012) which might be at play here where 'just in case' is a one-way conditional, which is what I suspect Rovane has in mind. This reading would mean the above quote would be claiming ($R \subset P$), that is to say, if something engages in agency regarding relations than that something is of the kind 'person'. This means that engaging in agency regarding relation is sufficient for personhood, but not necessary and therefore the concerning counter examples are no longer automatically excluded.

I could proceed if either of these readings of her text is correct. Advance directives are the provenance of human persons only so the concerns about non-human animals are not germane to the conversation. (Competent) persons with autism or other conditions, such as psychopathy, may create difficulty in recognizing agency are nonetheless capable of drafting advance directives and therefore de facto engaging in agency-regarding relations. I do think it is still worthwhile to address this ambiguity in anticipation of some objection based on the non-dynamic nature of isomorphic content.

[99] See Wade, 2007, De Waal, 2016 for a brief introduction into this issue. A video of De Waal's TED talk can be found at: https://www.youtube.com/watch?v=-dMoK48QGL8, (Timmango, 2016) which is both entertaining and enlightening.

If Rovane intended a biconditional statement, then to be a person requires active involvement with other agents. The extended mind thesis creates the possibility that a person can exist apart from a body/brain she actively controls. A biconditional reading would negate this possibility despite the purpose and intents of an advance directive. Advance directives are *for* influencing the behavior of others, but if Rovane's definition of person is biconditional, there would be no one to do the influencing, only words on paper with no moral force.

Change of Person- The Pleasantly Demented

There are rare cases[100] in which a patient has become demented but has not fallen into the awful state our intuitions seem to believe and imaginations default to when discussing such matters. In a widely cited narrative piece Andrew Firlik describes one such patient, Margo, as being a shadow of her former self yet nonetheless living a pleasant, yet limited life (Firlik, 1991). Margo spends her days happily rereading the same novels, visiting with people she may or may not remember and painting the same simple logo over and over again. Despite her illness, Margo does not appear to experience the suffering often associated with her condition. On the contrary, she is living a pleasant, simple life.

Let's assume for a moment that Margo had an advance directive which swore off lifesaving treatment should she find herself seriously demented (pleasantly or otherwise). At the time of the drafting she may have described the condition she would later find herself in as unbearable and certainly not something she would ever wish for herself. However, once she is in a demented state and has no biological memory of drafting an advance directive she may describe her life in positive terms given her current mental state.

[100] That is, rare relative to more common forms of dementia. "Pleasant dementia" is not a medical term or officially a diagnosis. As a result, quantifying the number of cases of this kind is a tricky, subjective business.

If the Margo who drafted the directive is the same person as the pleasantly demented patient (as the lucidity argument indicates) we can say that Margo has changed her mind about what condition under which life is worth living. In the previous section we applied the directive to a patient who was no longer a person because the directive is a persistent memory or act of will which was the most relevant piece of a person's mind that remains. Yet we should not outright dismiss Margo's contemporary wishes. This is a case of the extended mind causing extended cognitive dissidence. There are two legitimate, coexistent desires prescribing different actions to third parties. There is reason for respecting both the wishes currently residing in Margo's biological brain (to go about having an ignorantly blissful time), but also reason for giving weight to the desires she extended into the future in order to secure the end of life she then desired. The directive was made without knowledge of how the person would feel at the time the directive would become applicable (*Change of Heart*, one of the four KPADs). But this change of heart is not something unknown to us. Not only do competent patients change their minds about what treatment they find acceptable when actually faced with a medical condition, but also people are remarkably bad at choosing what is in their best interest (Dworkin, 1994, Velasquez, 2012).

We are not necessarily bound by our previous mental states be they performatives, beliefs or desires. We see this quite often as people's tastes change over the course of their lifetimes. Take my friend Wayne, for example. I have known Wayne for nearly thirty years and in that time I have never known him to like cheese. He did tell me once that when he was younger (before we met) he would often come home from elementary school and eat a Kraft single right out of the package as an after school snack. Today he shows revulsion even recalling the memory. Throughout the course of his life, Wayne once had the internal mental state of a person

who liked processed cheese and at a later time in his life, the opposite mental state. Wayne remembers today that he once held the belief that cheese was tasty and endorsed his eating it (tacitly, by doing do). While having this memory about this past life accessible to him, he also today believes (quite strongly) that cheese is not something he wishes to eat.[101] If Wayne's mental states are considered without temporal separation we would see a contradiction; Wayne both likes and does not like cheese. This, however, is not a contradiction, as the temporality is intrinsic to beliefs, memories and other kinds of mental states. Instead we say Wayne has had a change of heart and we rarely have a problem with people doing this sort of thing. We do not say Wayne is now 'wrong' for not liking cheese, but rather only that his tastes have changed and his outward expressions of those tastes reflects this.

This example relates to advance directives in the following way: say a person in good health believes being paralyzed from the neck down is not a life worth living and drafts an advance directive to that effect. She states that should she become paralyzed in this way, and is unable to give medical consent, no attempts should be made to save her life. Some years later she in in a car accident and is thusly paralyzed. She is conscious and competent to make medical decision and has been consenting to and accepting medical aid and assistance. After some time, a doctor comes to learn of her previous directive in which she forswore lifesaving treatment (which in her compromised state could be considered virtually any treatment or assistance). In light of the discovery of the advance directive, should the medical professionals now stop treatment despite her current acceptance of medical interventions? Of course not. The choice for the doctor as to which instructions to follow is an easy one.

[101] He did once argue for an exception: "Except on pizza. Only weirdos don't like pizza."

As agents we are allowed to change our minds. If a competent person writes an advance directive and later, as a competent patient, rescinds that directive or makes a choice not in accordance with that directive we allow her to do so. Though advance directives are externalized performative actions, they are not, actually or morally, unimpeachable by the drafter. We rightfully respect the contemporary wishes of the competent patient over her previous written directives. Its externality to the body of the person to whom it belongs does not lend any more permanence or deference if the patient is still a competent, active mental agent.

But the case in which the patient becomes pleasantly demented presents some challenges. A pleasantly demented patient is, by definition, not competent, but also by definition not experiencing suffering. What should be done in cases where the wishes of a pleasantly demented patient contradict the wishes of the formerly competent person she once was? It could be argued that in this case we do not have a single person who changed her mind over time, but instead a single biological animal with two different persons at different times, the drafter and the pleasantly demented individual. To complicate matters the past mental life has persisted into the present in the form of an advance directive. A person who is now is no longer in existence, or at least no longer experiencing life as she once was, has imposed her desire into the future. Is this really a conflict to be worried about?

Pleasantly Demented but Still the Same Person?

Earlier in this chapter I made the argument that a demented person was the same person over time so long as they continue the exhibition of traits indicative of personhood. The criteria of psychological continuity/connectedness of persons (a la Locke, Reid and Parfit) and my lucidity argument lead us to the conclusion that the pleasantly demented patient is the same person, albeit a person with different desires, goals and memories as the person she once was.

The conflict here is between a competently drafted directive (say, by Margo) and the incompetent contemporaneously expressed wishes[102] of the pleasantly demented patient (call her Margo*).[103]

The result is that we disregard the directive. This case is as simple as Margo changing her mind. What needs explanation here is the justification of how we come to favor the wishes of Margo* (who is, for practical purposes here, incompetent) over the competently drafted advance directive of Margo. Can the incompetent patient change the competent person's mind? Competency is not an all or nothing proposition. A person can be competent in some respects and not others. We need to take care to not paint on competency with too broad a brush or rub it away with too strong an abrasive. A dementia patient can be competent on some days and not others. She can be competent in some areas and incompetent in others. If an otherwise incompetent patient says the room she is in feels too cold and is requesting a blanket so she'll be more comfortable we have no cause to disregard her assessment of the situation just because we have deemed her incapable of making complex medical decisions.

The question is, "is Margo*'s life a life Margo would want to live?" In order to answer this question, we need to consider the strength of the positions the different players, Margo and Margo*, have in evaluating the situation. Margo* is currently experiencing her situation whereas Margo was speculating about what it would be like. Both Margo and Margo* think about Margo*'s situation from the position of how it would be for her, as she is now to find herself in such a state. The difference is that Margo has to do so from and outside perspective

[102] This does raise an interesting question (which I don't believe I have time for here) as to whether it can be incompetent to enjoy one's self.
[103] The titles of Margo and Margo* are meant to denote a difference in stages of the continued identity of a single person, not numerically different persons.

while Margo* has the advantage of actually experiencing the situation and knowing firsthand what it is like.

Imagine if Wayne, as an elementary school student decreed that Wayne* (Wayne today) must eat cheese every day because Wayne thought it was delicious and could not imagine ever changing his mind. Wayne* is in a far better position to determine what he likes and dislikes than Wayne ever could be through speculation alone. There was always a chance that Wayne* would like cheese just as Wayne did; perhaps even a strong inductive argument supporting this possibility. Nonetheless this speculation is never going to be as certain as firsthand experience in these matters. There is a dis-analogy which needs addressing here. Wayne has remained competent over the time in question (arguable he has become more competent as he was a child at the beginning of the story and an adult at the end) while Margo has lost competency as she progressed to being Margo*. For other purposes this discrepancy would certainly be relevant but here it is not so much a problem. Although the pleasantly demented patient may bear little in common with the competent drafter of an advance directive, the lucidity argument makes it the case that the patient remains the same person; as does Wayne. The difference in the expressed desires of the contemporary patient and the dictates of the directive are the result of a change of mind of the patient once she was able to experience the facts of her situation firsthand; which, again, is similar to the situation Wayne found himself in as time moved on. Though the pleasantly demented patient may not be an appropriate authority in matters concerning plans for the future or choosing between surgery or drug therapy for dealing with an acute medical issue, she remains authoritative on how she feels her life is going for her right now. The current experiences Margo* has now are acceptable, even desirable, to her whereas Margo believed they would be intolerable. She has changed her mind about the relative badness of the situation given

the change in her position relative to the situation. Just as Wayne* decided cheese was gross, Margo* decided her life was worth living.

There may be a pressing incongruity between Wayne* and Margo* situation found when we look at the experiential versus critical interests of each respectively. Wayne's earlier desire for processed cheese and his current revulsion of nearly all things cultured dairy are based on experiential interests. On the other hand, Margo's advance directive may have been (but not necessarily) concerned with critical interests.

What evidence do we have that Wayne's (experiential) interests have changed: his words and behavior. If Wayne says "I don't like cheese" it is not appropriate to reply "nope, you're wrong. You do like it."

Critical interests are less transient and more dearly held than experiential interests but are, nonetheless, able to change. This change might take a great deal of time, but it could happen over a short period as well. If Margo's critical interests have changed, we also must rely on her words and actions if we are to come to know a change has taken place. Margo's reason for drafting her advance directive was to avoid a situation she believed to be "unbearable and certainly not something she would ever wish for herself". Margo*'s experience of her current position stands in opposition to this previous belief. She is, as stated, "living a pleasant, yet limited life."

Margo is speculating on the experience of the badness of Margo*'s life and this badness did not materialize. Margo and Margo* are the same person and she has "decided" her directive incorrectly predicted how Margo would feel about the situation of Margo*. If Margo had retained competency up to the time when she would have otherwise become Margo*, we would not judge her desire to continue living as being in error. If Margo consciously and intentionally

choose to live a "simple, yet limited life" we would not insist that lifesaving treatment be withheld because her current assessment of her quality of life does not square with her predictions about her future self. Even if Margo's life, while competent, were filled with severe pain and limitations and she choose to receive lifesaving medical care we would not say she is incorrect in doing so (i.e.- that she is mistaken about what she desires). Her contemporary words and actions supplant those contrary words and behaviors in the past.

To put the matter another way; Star Trek V: The Final Frontier is considered by Star Trek fans, almost universally, to be the worst of the 10 films in the franchise. Though I am a dyed in the wool Trekkie, I like Star Trek V. I have had countless debates with friends and other fans about the merits of Star Trek V and have yet been able to adequately "defend" my position. I have never convinced any of the orthodox that Star Trek V is, at the very least, not as bad as they think. But my inability to sway the loyal does nothing to change the fact that I like Star Trek V, logic and argument be damned. This is in the same manner that Wayne cannot argue himself into liking cheese and Margo's argument for the position expressed in her advance directive does not change how Margo* now feels.

Granted, this exact argument only applies specifically to cases like Margo's where the critical interest can be interpreted as being about experiential interests, and this will not always be the case, as there are other reasons people choose to draft advance directives. What do we say of such cases as a critical interest to not have one's deeply held religious beliefs violated or a desire to see one's estate be used for charitable purposes rather than liquidated and spent on heroic medical interventions? Here the critical interest is not tied to, or a stand in for, an experiential interest in any obvious way. Are we to ignore the advanced expressed desire of our patient to have her critical interests protected just because she has become incompetent to protect

those interests in real time? In a word, yes, but how we put this reasoning into action may get more complicated.

Critical interest, like experiential interests, can change. Likewise, it is only through an examination of the words and behavior of the interest haver that we are able to, from an outside point of view, determine when this is the case. A pleasantly demented patient meets the criteria of being the same person over time as she was before the dementia set in (per the lucidity argument) and it is the prerogative of that person to decide and let it be known when her interests have changed. If a demented patient is expressing different critical interests than she did previously, we are under obligation to disregard the pervious wishes, but sometimes we, and the patient, can have it both ways.

We can see the expressed wishes of the demented patient and the wishes of the directive as being a kind of cognitive dissidence. The patient believes two (or perhaps more) things without being aware of it. This unconscious dissidence is not uncommon even in healthy persons. Let's say I believe there are eight planets in our solar system (because I believe Pluto should no longer be considered a planet). Assume also that I have the dispositional memory that the mnemonic "My Very Excellent Mother Just Served Us Nine Pizzas"[104] is a good tool for remembering the names of all, and only, the planets orbiting Sol. So long as the mnemonic is not called or otherwise brought into my conscious mind (at which time I can make changes to my beliefs to make them self-consistent) I, in fact, believe two things which are not logically possible to believe together. If Margo had a desire to not see her religious beliefs violated (which she made known through her directive) and Margo* is consenting to treatment which is opposition to this critical interest, we are faced with the same kind of scenario. An important

[104] This mnemonic will come up again later in slightly different context.

difference being that it is not (or at least may not be) possible for Margo* to have the beliefs of Margo brought into very consciousness. For Margo* the advance directive will always remain a dispositional memory.

Because of the sameness of person, the desires expressed by Margo* ultimately rule the day. However, Margo* has been deemed incompetent, so to at least some degree we don't have to do *exactly* what she wants. I love my son, but he does not always think things through and make the best choices with the information at his disposal; which is not surprising, since he's only two years old. Being two, he is incompetent in many ways, just as Margo*. If my son wants to play outside he is expressing an experiential interest in doing what he believes will lead to having a good time. Under normal circumstances there is nothing wrong with granting Henry's request. However, if there is a thunderstorm, tornado or hail storm, allowing him to go outside would lead to disaster. But try convincing him of that. Whether or not Henry knows it, he has a critical interest in being kept safe. There is a conflict between his expressed desire to play and his (unknown to him) critical interest in remaining safe. Despite his protestations, Henry will be kept inside until such time it is safe to go out and play. Though this way of thinking may not work with all critical interests, we can at least *sometimes* have it both ways with the demented, religious patient. If Margo was an observant Muslim she would not wish to eat between sun up and sun down during Ramadan. If, however, during the holy month Margo* requested some Cheez-its[105] and apple juice around noon we would have a conflict in need of resolving. Asking Margo to forgo the snack or distracting her until sundown and giving her a treat at a religiously appropriate time satisfies both the critical interest in the advance directive

[105] Though sources differ slightly depending on interpretation of the relevant text and strictness of the practitioner Cheez-it are, to my surprise, generally considered to be halal.

and the experiential interest of eating the desired crackers. Margo*'s incompetency to make certain decisions means that others are in position to take steps to satisfy both sets of conflicting desires to whatever degree is possible. In the end, however, if such a middle ground cannot reasonably be found, the lucidity argument and the above application make it the case that what Margo* says goes.

Change of Person- The Severely Demented

In the far more common cases where the patient is not pleasantly demented, but rather just demented, the advance directive take on a different role. When the capacity to competently change one's mind is absent, and the directive may be the only remaining, coherent mental state attributable to the patient. Patients are sometimes rendered into a state such that they are no longer a person. This is most obviously the case when a patient is dead or has entered an irreversible PVS, but is also possible if dementia proceeds to a far enough degree and yet some level of consciousness remains.

In this section we will discuss cases in which the patient is a person (what happens when the patient ceases to be a person was discussed in a previous section). She can express desires, communicate with others and perhaps, from time to time, experience the occasional pleasure. Since the patient has the ability to communicate and express desires, there exists the possibility her expressed desires might differ from those spelled out in her directive. But, they might be in agreement as well. Expressed wishes and written directives can be quite nuanced and highly specific to the patient. Patients do not always express an all or nothing approach to their care. Not every time it becomes necessary to refer to a patient's directive is it a matter of life and death. Sometimes it is a matter of trying to determine what therapeutic option(s) a patient might choose as a means of treatment. Is the surgeon tasked with removing a cancerous tumor

authorized to amputate, or must she wait to wake the patient up and get further consent once the full extent of the cancer's spread is known? Since it would be impossible to discuss all possible ins and outs and conflicts between advance directives and the currently expressed wishes of the patient, I am going to limit the conversation here to questions regarding life-sustaining/saving treatment and whether the directive and the current patient either requests or denies such treatment. This does greatly limit the scope of what we can talk about here, but I believe that this theoretical overview can be applied with specific cases where the stakes are not so high. There are four relationship kinds between demented patients and the dictates of the directives previously drafted. The patient can express a desire to live or not and the directive could express a desire to be kept alive or not. Form this we can graph a simple matrix:

	Patient desires to live	Patient desires to die
Directive expresses desire to live	Agreement to live	Conflict (Type1)
Directive expresses desire to die	Conflict (Type2)	Agreement to die

Figure 2- Matrix displaying patient's current expressed wishes and those in patient's advance directive[106]

[106] The term "desire to die" is not the most elegant way to phrase the expressed wishes of patients in these cases. It would be more apt to say "a desire to see an end to my suffering, even if that means my death", but this is long and clumsy.

Half of the boxes have the patient and the directive being in agreement, but this does not mean conflicts do not arise in these cases. In instances where the directive and the expressed wishes of the demented patient agree we will, of course, uphold the directive,[107] however, depending on our reasons, doing so will either add strength or create weakness in our justification.

Each of these four scenarios needs to be investigated separately. Cases in which the incompetent patient agrees with the dictates of her directive (either to live or die) are somewhat easier to handle and do not require the same amount of legwork.

I previously argued that an advance directive is part of a person's (extended) mind. This being the case, we should not view the directive and the patient as separate entities. The advance directive is as much a part of the patient's mental life as the experiences she is presently having.

I add this reminder as a means of lightening my argumentative load for the next few sections. Rather than repeat, over and over again that the contents of an advance directive are the contents of a patient's mind, I will refer to the 'continued extension argument' or the CEA. The continued extension argument states that those wishes expressed in the advance directive are just as strong as they would be were they uttered at that moment, because, in reality, they are. Those wishes are the end result of a performative act. In this case they are uttered through a medium other than the patient's literal voice. When a patient becomes incompetent to make her

[107] I am not assuming the competent patient's wishes are sacrosanct and not immune to overruling by interested third parties. There are a very small number of cases where it may be morally prudent to ignore the patient's subjective desires so as to act in the objective best interests of the patient. If such cases arise while considering the implication of the extended mind the same arguments for and against a disregard for the patient's autonomy would hold. (see Arras, 1988, Varelius, 2006, Veatch, 2000)

own health care choices known we should not forget the directive is an active part of her mind. The directive is a persistent mental state or act of will which has been expressed by the patient.

I will begin by considering the cells of the matrix in which the directive and the current wishes of the patient are in agreement before moving onto the more difficult issues involving disagreement. After the resolutions to these conflicts have been considered I will give an examination of the results. I believe this examination will further justify the way in which I resolved the conflicts between the current patient and her directives.

Agreement to Live

In the end we will uphold the wishes of the patient both past and present and help the patient continue to live and make her life as pleasant as possible. The directive is constant here, as argued by the CEA; the patient, as she finds herself now, is the wild card.

A desire to live can be expressed in a few different ways. One way is the earnest desire to continue living and the desire to see that life is as pleasant as possible, such as would be expressed in an advance directive or by a competent patient making her own decision known as the need arises. Another form is the "desire" to live which comes in the form of an incompetent patient refusing medical treatment she, erroneously, believes is harmful, painful or unnecessary for her. A third case, likewise to the second, an incompetent patient can request treatment she believes is good for her but in fact would be harmful or even fatal. In the second and third case an expression to live but is made on incompetent grounds.[108] If I am demented, yet desire to live, I might think the nurses are trying to inject me with bleach rather than antibiotics. In my desire to not die by poisoning I refuse treatment. I have expressed the will to live but have done

[108] The second and third forms are two sides of the same coin. There will not be much need to talk about them as if they are different. Parallel reasoning applies in these cases.

so, tacitly, by refusing the treatment which will help keep me alive because I am not in a frame of mind capable of making such self-interested choices.

The relevant issue here is what interests are being defended by the patient. The expressed desire to live, both in an advance directive or in an incompetent state is an act directed toward the protection of a critical interest: life itself. Actions taken by medical personnel aimed at protecting this critical interest are done in the patient's best interest (i.e. – in accordance with her critical interests) whether she is aware of this fact or not. The patient in the past, through the continued mental state embodied in her advance directive, and the incompetent patient now through her protestations are in agreement as to what is the desired end result of medical treatment, to live.

The directive expresses a desire to live and the patient is refusing sound treatment on the basis that it will cause her death. What's in question is not the patient's competency, rather her autonomy. The directive is an external manifestation of a competent mental state, the demented refusal of an injection of antibiotics is an incompetent refusal of medically necessary treatment. If we administer treatment we end up respecting the patient's wishes (embodied in the advance directive) in the first case we respected the person's wishes (giving them the necessary shot) but in the second we violate her contemporarily expressed wishes, despite giving the patient her desired result. If we are acting in accordance to the directive we are being more respectful of the person's wishes, despite her protestations to the contrary.

The advance directive is a performative act and it may in fact be the last known, or existent, competent mental state of the patient. We have a choice between the competently expressed wishes of the performative act versus the incompetent and delusional reactions of a very ill person. Both of these mental states are mental states of that person (via the CAE and the

application of the lucidity argument). If we had a patient who was drifting in and out of competence, perfectly lucid one moment and conscious, yet incompetent the next, we would defer to the wishes of the patient while lucid. With the advance directive we have much the same thing. All that has changed by using an advance directive is the location and medium of the memory/desire and thus how it is becoming known or reported to outside parties.

This argument can also be used in conjunction with Ronald Dworkin's notion of critical interests. Advance directives are, at their absolute best, well thought out, considered decisions of the drafter. It is a type of document that requires the upmost competence and highest degree of autonomy to draft legitimately. As such it is easy to see how we are quite safe in considering the values interests expressed in these documents as those most dear to the drafter. A delusional patient cannot make the same claim of thoughtfulness.

If the patient's dementia is not too far progressed and she does have moments of clarity, if not competency, the lucidity argument applies. Even if the patient's "old self" hasn't been seen in a while we cannot (logically) argue with certainty that she will never resurface. However, it is reasonable to believe, and act upon the belief, that if the "old self" has not been seen for some time it is *most likely* not going to happen. To argue otherwise is, essentially, akin to "waiting for a miracle" as a medical treatment. Perhaps it is logically possible, but logical possibility doesn't make it a prudent, or reasonable medical prescription.

There is, however, a transitive argument that does apply in such cases. The currently expressed desire to live is transitively related to the desire to live expressed in the directive. If the patient has, through one means or another, competently or not, consistently expressed a desire to live, she has done so using the same neural hardware both in the past and at present.

These desires come from the same place (both literally and metaphorically), though the 'arguments' for what will actually keep her alive may be wildly different.

If a patient is refusing antibiotics for fear of being injected with bleach, her present protestations should be ignored in the service of her desire to live for three reasons.

1. The advanced directive is a mental state of the demented patient, though one she cannot herself access in her current condition. It is a dispositional act of will which she may not be able to access, yet others can. Despite not being able to access the memory or act of will, it is still her memory or act of will. (Much like a memory I know I have but cannot call to consciousness despite making the effort).
2. The same interest is being argued for by both the current patient and the drafter of the directive, who are, as previously, the same person.
3. Failing that, the lucidity argument and its sub-conclusion regarding sameness of neural hardware indicate a degree of psychological connectedness between the two.[109]

By treating the patient against her currently expressed wishes we are actually assisting her in achieving her competently and incompetently expressed goals and giving full consideration to the mind of the patient both internal and extended.[110]

[109] I am not here going to give arguments concerning the relative strengths of instance of psychological connectedness. It is enough to rely on Parfit's previous discussed distinction between psychological connectedness and (mere) psychological continuity and his conclusion that connectedness is more meaningful for personal identity than continuity.
[110] This is similar to arguments made by Varelius (2006) when discussing overruling a competent patient's expressed wishes because if doing so respects the autonomy of the person as a whole rather than just the autonomy of the person to make this one decision.

Agreement to Die

The reasoning behind this decision closely parallels to the previous case in which the patient and the directive expressed a desire to live. By withholding treatment, we are giving the patient (both past and present) what she wants, an end to her suffering sooner rather than later, whether she realizes it or not. If we uphold this wish as a result of respecting her advance directive, we have given preference to her competent wishes rather than incompetent misunderstanding of her situation.

Just as with an agreement to live, there are different forms in which a desire to die could be expressed depending on the competency of the patient. There is a difference between a competent desire to end one's suffering and incompetence leading to a misunderstanding of what benefits the patient stands to gain from further treatment.

In this case we'll imagine a patient who is slowly dying due to a painful, physically and mentally degenerative condition. Her advance directive states that should she find herself in such a situation and it is the considered opinion of her caregivers that her case is irreversible then she should not be given life-sustaining/saving treatment for any other conditions which might end her life sooner rather than later than her underlying illness.

After losing competency the patient contracts bacterial pneumonia. The patient, now incompetent, and experiencing the distress and pain that comes from a compromised ability to draw a full breath, demands an injection of antibiotics which she believes would be lethal to her (maybe believing 'antibiotic" means "kills all life") but in fact, would cure her pneumonia and result in her living longer than her advance directive expressed a desire for.

There is an understandable desire to end the patient's distress by administering the antibiotics. Yet, doing so would only prolong a different kind of suffering, which is exactly what

the advanced directive was intent on preventing. Again, we are faced with a competent directive and a demented expressed desire both aimed at the same end. Ultimately treatment for the pneumonia should be withheld in accordance with the advance directive.

Conflicts

Far more difficult cases concern instances where the directive and the currently expressed wishes of the patient disagree on an acceptable course of treatment; one expresses a desire for lifesaving/sustaining[111] treatment while the other gives directions for such treatment to be withheld. In the above matrix I labeled these disagreements thusly:

Conflict (Type 1)

The directive expresses a desire to receive lifesaving treatment, while the patient currently expresses a desire to die rather than continue in her current state.

Conflict (Type 2)

The directive expresses a desire for the patient's life to be allowed to end as soon as possible, even if this means withholding sound medical treatment, should certain conditions be obtained, while the patient currently expresses a desire to continue living.

Unlike the cases of agreement above there exists here a genuine uncertainly as to how to proceed. In fact, good arguments exist on all sides (which were previously citied throughout this dissertation). In cases of Conflict Type 1 we have reason to adhere to the directive as well as those wishes the patient is currently expressing. Ditto for cases of Conflict Type 2. This being the case I will give summary arguments why we should or should not adhere to the advance

[111] Henceforth, unless otherwise noted, "lifesaving treatment" refers to both lifesaving and life-sustaining treatment.

directive in the particular kind of conflict and then argue for which is to be preferred in light of the extended mind.

In both types of conflict between advance directives and the contemporarily expressed wishes of the incompetent patient I will give brief arguments for listening to the dictates of each (additional details of these positions will be given, as needed, in the next step). I will then give additional arguments in favor of one over the other though the same resources (the CAE and the lucidity argument) figure into resolving both types of conflict the reasoning differs.

Conflict Type 1

The Advance directive expresses a desire to live while the patient currently expresses a desire to die.

Argument for following the advance directive

The continued extension argument is the obvious starting point here. There is the competently expressed desire in the advance directive in which the patient requests reasonable steps be taken to preserve her life as long as possible, weighed against the incompetently expressed desire to die so as to end what the patient perceives as suffering. All things being equal, we should favor the competent over the incompetent.

The patient is currently expressing a desire to live through the advance directive. This expression was initially made when the patient was competent and through the lucidity argument that patient still exists and has existed throughout the entire progression of her condition. However, she is also currently expressing a desire to have her life end. The difference here is that the advance directive, if it is valid, is a considered opinion and not one based on the, perhaps short term, unpleasantness of the situation she now finds herself in. When drafting the directive the patient (if competent and properly informed) would have known this sort of dissention would be possible, perhaps even likely. The drafting of the directive shows a concern for her critical

interests (of continued living) over experiential interests (avoiding unpleasantness and suffering). The continued extension argument demonstrates the importance of upholding the advance directive as a means of defending the patient's critical interests, the reason for the drafting of the directive in the first place.

Under this reasoning the advance directive should be upheld and the patient's life sustained.

Argument for listening to the patient

We might argue that the incompetent patient is in some ways more competent (or at least more informed) given her actually experiencing the situation which the competent patient only encountered theoretically. Levels of pain can vary greatly. Perhaps the level of pain the patient is now in would be something completely tolerable were she competent, but in her current state even relatively minor pain seems intolerable, a judgement the patient is perfectly competent to make. Being able to understand one's suffering or pain may lessen the unpleasantness of it. Consider the different reactions I have to getting a shot from my doctor versus the reaction my son has. There is (I don't believe) a significant difference in the actual amount of pain involved, however, my competent understanding of the situation mitigates the pain, whereas my son has no context. For him, the shot causes pain for pain's sake.

Of course, we make the argument that we do not allow incompetent children to forego shots because they find them unpleasant because we have the child's (hopefully long) future to consider.

The critical interest of having one's life last as long as possible fails to out outweigh the patient's experimental interests for two reasons. First, there is no appreciable future to consider. There is not a long life left to consider so there is not much of that interest to be protected. In the

case of the terminally ill, incompetent patient no such future is possible. The pain being experienced by the terminally ill, incompetent patient is not aimed at some long-term greater good or possible far future benefit. This is merely a condition the patient must endure in order to continue living and perhaps this condition constitutes the entirety of the patient's remaining life. Additionally, the combination of limited time the patient has and the condition she is in make it the case there is little that the patient can do to define her personhood or meet goals which are important to her (this argument is borrowed, heavily, from Gill, 2005).

Secondly, what life does remain, on its face, is not one that perhaps many would find worth living. There is something objectively bad about living in a condition of constant pain despite what the drafter of the directive thought. As the patient suffers through the end stages of her disease to sustain her life, regardless of her protestations to the contrary, is forcing the will of a past, ignorant person onto another; a situation we justifiably do not generally allow.

Which Wins?

In cases of Conflict Type 1 we should uphold the advance directive and sustain the patient's life. What has not yet been discussed are the semantics that led to the drafting of the directive. Though the drafter was not experiencing the unpleasantness of the end of her life firsthand, when drafting the directive she was not completely unaware of the possibilities. The directive was written to protect the critical interest of sustaining life and a competent drafter would know that the possibility exists that there might someday be very little (temporal) life to protect, however, it was important to the drafter that it be protected nonetheless.

What are the reasons a person would want to hold onto a life that perhaps most people would never want to consider living? Some reasons might include:

- A religious belief that one should do whatever is possible to live as long as possible.

- A belief that one learns something important about the human condition through the suffering at the end of life.
- A belief that "it's my life and I want as much of it as possible."

The reader might find one, some or none of these[112] as a compelling reason to continue living, however, perhaps the drafter of the directive did. One, or all of these motivations (or others) were the impetus of the drafting of the directive and the length of time these interests could be upheld is not relevant to the one whose interests are in question (unless length of time is something spelled out in the directive, which does not change the argument here. It just shifts the question over to Conflict Type 2).

It almost seems inevitable that a conversation about advance directives will at some point turn to Odysseus and the sirens as an analogy.[113] After very briefly summarizing the story I will apply an analysis in consideration of the extended mind thesis. Lastly, I will move the conversation back to conflicts with advance directives, which, having Odysseus's tale under our belt, should be easier to get a grasp on.

While off adventuring after the Trojan War, Odysseus and his crew need to pass through a perilous stretch of water surrounded by the manipulative sirens. The siren's song lead sailors astray, causing them to crash on the rocks and consequently sink. Odysseus orders his men to plug their ears with wax to avoid the danger. Odysseus himself wishes to hear the siren song so has himself tied to the mast of his ship to prevent himself from steering the vessel off course. He

[112] There are others, of course, that the reader may or not find equally convincing or lacking.
[113] Instances of this are far too numerous to list. I will say that I first encountered this while attending the dissertation defense of a graduate school colleague, Jenny Swindell (now Dr. Jennifer Blumenthal-Barby at Rice University).

tells his men to not let him down, no matter how hard he protests. In fact, if he protests they are to tighten his bonds.

What are the sailors to do once Odysseus begins to cry for release? Odysseus has previously expressed his desire to not be released, but while the sirens are singing appears to have changed his mind. Do the sailors follow the previous expressed order or do they abide by his currently expressed wishes to be unbound? In the epic the seamen follow Odysseus's advanced instructions and leave him tied to the mast, which, it turns out, is what the extended mind thesis would have them do.

Odysseus valued the safety of his person, crew and ship and took the necessary steps to see that this interest was taken care of. Odysseus, in anticipation of his incapacitation, eternalized his desires for action into the minds of his crew. Knowing that his mind would be incapable of carrying out his own wishes, Odysseus made a copy of his desires (in the brains of the crewmen with wax-filled ears) which would not fall victim to the same coercive factors. The issuing of his instructions was a performative act by which Odysseus said, 'I hereby charge you with carrying out my wishes.' The content of these wishes is the same whether in Odysseus's mind or temporarily in the minds of his crew. If Odysseus was of his right mind he would have carried out the instructions himself. The extension of his wishes ensured that he was going to be in his right mind when it appeared otherwise. The extension of this desire into the minds of the crew (which for this singular purpose were now a part of his mind) ensured that his competent choice to remain tied to the mast remained an active cognitive state, just not a cognitive state within his own brain.

Let's consider an analogy within an analogy. Odysseus is back on his home on the island of Ithaca and is building a storage shed with his son, Telemachus. To place a ceiling joist

Odysseus needs to hold it over his head while Telemachus drives in the pegs to hold it in place. Holding the joist aloft is not easy. It is heavy and Odysseus needs to contort himself to get it into just the right position. His body is screaming at him to let go and end the discomfort his body is going through. If he lets go his son will fall and injury himself and the shed will not be completed. Odysseus knows the pain is temporary and pushes through because he has a desire to see his son safe and the structure completed. Throughout the construction project Odysseus has remained 'in his right mind' and suffered through his situation despite his experience telling him to quit.

While the sirens are singing Odysseus does not have (on his own), the cognitive resources to see the situation through, so he borrows resources from others to keep him 'in his right mind.' The part of his mind that 'pushes through' his current conflict with the sirens has been externalized. He cannot, internally, choose to ignore the part of him crying to be released, so the crew is performing this cognitive work. They are not doing so *for* him, but rather *as* him. The overall analogy should be fairly clear. Odysseus's request to remain tied to the mast is an advance directive and his pleas to be released are analogous to the expressed wishes of the incompetent patient. Just as Odysseus issued instructions to ensure his wishes were carried out, so too does the patient through issuing an advance directive. The directive is not just guidance or a suggestion by the patient. The performative act of issuing a directive is done so to ensure that those competent desires of the patient remain an active part of the patient's mental life. The directive is not a document to be accessed in the absence of the patient's input. It *is* the active cognitive input of the patient. When a directive expressing a desire for lifesaving treatment conflicts with the contemporarily expressed desires for treatment to stop, we are not witnessing a change of heart on the part of the patient. We are seeing an internal struggle become

externalized and therefore observable. If a patient were able to hold both mental states (the desire to sustain life and the desire to stop treatment) the conflict would be internalized and there would be no question whether or not the conflict was mental; it obviously would be. But, given that the patient cannot sort this out internally, the task falls to outside minds which have been conscripted into the patient's mental processes. Both the current internal desires of the patient being expressed, and the written directive are fully mental states of the patient. The choice we face is between a competent and incompetent mental state. The directive is competent and therefore should be favored and carried out. Lifesaving treatment should be given.[114]

Conflict Type 2

The Advance directive expresses a die to live while the patient currently expresses a desire to live.

The arguments here are in some ways similar to those found in conflict type 1, and parallel reasoning applies in sorting out this conflict but since we have a patient expressing a desire to live despite what was said in her advance directive we may find ourselves dealing with some ethical squeamishness. We may be loath to say to a patient, incompetent or not, "I know you *say* you want to live but you said before you would rather die, so I'm afraid we're going to have to let you go." I will take the same initial approach and give outline of arguments for listening to the directive and the patient after which I will delve more deeply into the argument in favor of the patient and show why this reasoning is flawed. In the end I will conclude that the directive should be followed, all things being equal, but there may be times when the incompetent patient may override this pervious decision.

[114] This is to say there are times when it is right to ignore the advance directive and choose another course of action, even one contrary to the directive. This is covered in chapter 4.

Argument for following the advance directive

The continued extension argument here upholds a patient's autonomy to decide when she should die. When the directive was drafted the patient had a set of beliefs about what kind of life was or was not worth living. The patient still has these beliefs, via the continued extension argument, despite her currently expressed disagreement. The patient is expressing a desire to both live and not live in her current condition. Since she cannot have it both ways, we should turn to the choice that was made when the patient was competent (relative to her current state).

The possibility exists, and is known to the drafter, that when the spelled-out circumstances come to pass she might not find them as intolerable as she suspected. If she is competent when faced with a terminal diagnosis she would need only say the word and her directive would be ignored. But advance directives are not written for when we are competent and able to express our wishes. Whatever the metaphysical status of the future patient is, the drafter does not identify herself with that person; the drafter believes she would no longer exist despite the continuation of a biological life. She believes it is preferable that this biological life, her biological life, to end as closely as possible to the ending of her ethical and metaphysical life.

Argument for listening to the patient

Though the patient has been deemed incompetent we should not be so quick to dismiss an expressed desire to live. Whatever goods the patient believes (incompetently or not) her life holds for her disappear when she dies.

If the drafter of the directive no longer exists after the onset of incompetence (a point this dissertation disagrees with), then there is no reason at all to apply the directive. Since the person to whom the directive applies does not exist, the directive is moot. This is not an argument we will consider any further. What are we to do given that, according to the lucidity argument, the

drafter of the directive still exists and, via the continued extension argument, still desires for her biological life to end?

Similar to Margo and Margo*, the patient has changed her mind as to what kind of a life is worth living or what is a good reason to continue living. The current patient has changed the drafters mind even though the drafter is unaware of this fact. Whereas before the drafter believed life would not hold any (or at least not enough) good to continue, today the patient has found reason to think otherwise. The critical interests of the patient have changed in light of new information (i.e. – the information gained by being in the circumstances in question rather than speculating on them). The concerns which are manifest in the directive have not really come to fruition, in that the situation is not being perceived as bad either experientially or perhaps critically. Therefore, the current wishes of the patient are to be given precedence.

Which Wins?

By the end of this sections I will conclude the directive comes out on top, at least most of the time. But to make this argument let's first consider a further argument in favor of listening to the patient. assume that in this case the currently expressed wishes of the patient are our guiding principle. The argument is like that made in the cases of the pleasantly demented patient, but here pleasantness (or relative lack of badness) is not necessarily part of the equation.

Whatever remains of the drafter that continues to exist is perhaps little; the biological strata, some dispositional and semantic memories but not much else. The incompetent, yet life-desiring, patient has found good in life or at least value in living that she did not foresee when she drafted her directive. The semantics leading to the drafting of the directive were concerned with what the drafter believed, at the time, were unpleasant circumstances to be avoided or those not aligned with her (then) critical interests. It was not until these circumstances arose that she

was able to make (more) informed judgement about them. Granted, the patient is currently incompetent, but even Margo* was able to make some choices for herself, especially concerning what sorts of things make her life enjoyable.

This argument not only applies when the patient continued life is pleasant. The argument still holds in cases where the patient is suffering, yet still wishing to live. The patient may only wish to continue living because she believes death would be worse and give no further argument. The incompetent patient has found something good to hold onto, even if that something is the benefit of delaying the inevitable. Let's consider, again, Odysseus, taking several liberties with the situation.

Ten years of war and 10 years at sea is exhausting and Odysseus is tired of wandering the known world. Having narrowly escaped death on several occasions he's had enough. When he learns of the sirens and how listening to their song would end in shipwreck and death he decides this is as good a time as any to end this otherwise never-ending voyage. Odysseus instructs his crew that, no matter what, don't stop me from sailing this ship as the sirens call. I know it will kill us all, but it's what I want. "Besides" he says, "there's probably something on the other side of the rocks that is going to kill us anyway."

As they come into range of the sirens Odysseus hear music of a beauty like none he has ever heard before and turns the ship toward the rocks to get a better listen. He soon realizes that if he crashes he will no longer hear the song. He wants to get closer to the song and hear it for as long as possible. He asks the crew to please tie him to the mast of the ship so he can continue to listen.

The crew, unsure of what to do, reminds Odysseus of his request. "I know what I said," he replies, "but we can always come back and crash onto the rocks tomorrow."

The patient is having an externalized conflict between her current desire to live and her previous expressed (yet still active mental state) desire to not live under specified conditions. Disregarding the advance directive (for the remaining life of the patient or just temporarily) respects the patient's currently expressed desire without closing off the possibility of the directive being carried out at some time in the future. By ignoring the directive, we can allow the patient to experience whatever it is that she currently believes is good about the life she is living or maintain interests she now believes to be critical to her living a meaningful life. If, at a future time, her wish is to have life sustaining treatment withheld we can revert to the section of the previous matrix labeled 'agreement to die' and carry out her wishes in accordance with that argument.

There are compelling elements to this argument, but it is failing to consider a basic premise of how the extended mind thesis has been interpreted as applying in these cases. The advance directive is a mental state of the (otherwise) incompetent patient and, as has been said before, may be the last remaining competent mental state belonging to the patient. As those tasked with enforcing, or not enforcing, the directive we must choose between a desire projected competently into the future and the incompetent directions being given by the patient which are counter to the critical interests she drafted the directive to protect. If we take am impersonal view of the situation it should seem an easy choice to make; we do listen to a competent person before we listen to an incompetent person. But this is not an impersonal matter.

That ethical squeamishness I mentioned earlier is what makes this choice more difficult. I don't believe many people would relish the idea of saying to a person who could be kept alive, and says she want to remain alive, nope, you must die. A further argument is needed to make us more at ease with our discomfort.

Let's consider a case with some lower, or at least less immediate, stakes. Fred loves pistachio ice cream. The taste, the color, the smell, it is all makes eating pistachio ice cream a sublime experience. After a life time over overdoing it with the pistachio ice cream Fred has been told by his doctor he needs to lose a fair amount of weight else he faces a premature death. Fred has both the critical interest in going on living as long as possible (so long as life is at least reasonably pleasant) but the experience (experiential interest) of eating pistachio ice cream is one of the things that makes his life worth living. Not having an infinite supply of willpower Fred asks me to help him stay on the wagon and to not let him eat pistachio ice cream. Even if he insists I am to stop him, maybe even forcefully, from indulging himself.

Following the above logic in arguing in favor of the expressed desires of the incompetent patient (and Odysseus' desire to listen to the sirens) the exchange between Fred and me would go like this:

Fred: Michael, no matter what I say, do not let me eat pistachio ice cream.

Me: Sure, thing Fred.

(some time goes by)

Fred: Michael, I want to eat pistachio ice cream.

Me: OK. You can always not eat ice cream tomorrow.[115]

I have done the "right" thing insofar as I helped or did not impede someone from fulfilling an experiential interest which is harmful to no one but in doing so I completely disregarded to critical interest Fred's request was actually aimed at protecting, extending his life.

In instance of Conflict Type 2 listening to the patient rather than the directive is doing exactly this, preferring the experimental interest over the critical interest. Such a preference is

[115] Worse yet, I could have said "I'll get you a spoon".

sometimes permissible, perhaps morally incumbent, Conflict Type 2 is not a priori or even obviously such a case.

An advance directive, like Odysseus having himself tied to the mast, is an effort, in part, for the patient to protect herself from her future self should she begin to act contrary to her deeply held critical interests. None of this is to say that an incompetent patient can never countermand her advance directive. Though it seems a cope out to say we need to take these things on a case by case basis that is the most prudent approach so long as we begin with the caveat that the default position is to listen to the directive and only disregard it if there is a compelling reason to do so. This approach would be similar to those rare instances when we over turn the wishes of a competent patient. The onus is on those wishing to overturn the directive to show doing so protects or improves some critical interest of the patient or that the experiential interest has become more meaningful to the overall life of the incompetent patient.

Conclusion

The chart below summarizes what the extended mind thesis says we should do when considering an advance directive versus the expressed wishes of an incompetent patient:

	Patient desires to live	Patient desires to die
Directive expresses desire to live	Patient's life is preserved because of the continued extension argument and the patient's desire to protect critical interests.	Patient's life is preserved because of the continued extension argument and the semantics for drafting a directive which preserved life.
Directive expresses desire to die	Patient's life is allowed to end because of the continued extension argument and the patient's desire to protect critical interests.	Patient's life is allowed to end because of the continued extension argument and a desire to protect experiential interests

Figure 3- Matrix displaying outcomes of conflicts between a patient's current expressed wishes and those expressed in patient's advance directive

All things being equal we should respect the directive in all cases. When the patient and the directive agree our reasons for doing so are straight forward. When disagreement arises the continued extension argument coupled with considerations for the critical interest of the patient give us strong, though not inviolable, reasons to consider the directive as authoritative. Throughout this chapter I have considered the implications of the extended mind for advance directives and the KPADs related to personhood. The concern raised by some (Dresser, Buford, Bonner) is that the patient's suffering from dementia may not be the metaphysical same as the person who drafted the directive. An argument was made (the lucidity argument) that indeed there is a real sense in which these are the same persons based on the sameness of the neural hardware and the transitive nature of mental states.

Given this sameness of persons and the plausibility of the extended mind thesis, I articulated arguments for how we should proceed in cases where a patient has an advance directive and has become demented, or otherwise incompetent. In those rare cases described as pleasantly demented we ignore directives which forgo life-sustaining treatment. Though incompetent in many respects the pleasantly demented patient is still able to make some choices for her life and her tacit expression to continue living (by appearing to enjoy her life) amounts to a change of mind on the part of the drafter. In cases where the patient is otherwise demented we had to consider several factors when determining a course of action. First, is there a conflict between the directive and the current expressed interests of the patient? If there is no conflict the directive is to be followed because it is the continued, isomorphic mental state of the patient. If there is a conflict we choose the course of action which best respected the overall mental life of the patient both past and present,

Throughout this chapter there has been, admittedly, some over simplification of advance directives and how we implement them. Directives are not always as simple as Odysseus's orders to his crew (don't untie me, no matter what!). Language by its nature is ambiguous and there will be times when exactly what an advance directive is telling us to do will be unclear. We may have to choose between two, or more, seemingly qualified potentially proxy decision makers. Medical circumstances may have changed in ways the patient could not have anticipated necessitating changes to advance directive instructions. Further, the extended mind thesis may make it the case that a patient could change her mind and not even know it.

These problems of interpretation are the subject of the following chapter.

Chapter 4- Problems of Interpretation

The hope is advance directives would eliminate the need to make decisions for others. The very idea behind such a directive is to make the wishes of the patient known should she be unable to express them. In practice we have not eliminated the need to make decisions for others. There are primarily three reasons advance directives need to be interpreted. Some directives are vague, some become outdated in light of medical advances and others are made obsolete because of the changing wishes of the drafter. These causes for concern stem from two sources.

One is the inadequacy of language. Language, by its nature, is vague and paints an incomplete picture of the mind of the speaker. Even a simple request such as "I would like a steak"[116] while dining out carries in it several assumptions. It is assumed you want the steak in a few minutes, not several days. It assumes you want it on a plate and not shoved in your ear. [117] As we learn language we also learn the unspoken assumptions implicit in our speech. In our day-to-day lives this rarely causes problems and when it does we can quickly sort them out with a few follow-up questions. When the language at issue concerns the contents of an advance directives a mistake in interpretation can be the difference between respecting an individual's most deeply held wishes or not. Additionally, these errors can literally be the difference between life and death.

The second source of misunderstanding comes from a failure to appreciate the kind of performative action an advance directive is. As has been discussed (and argued for), an advance directive is a persistent act of will of the patient. It is not merely a suggestion or guidance. It is

[116] This exact example is borrowed from Searle, 1992.
[117] Vagueness is not always bad. There is such a thing as "useful vagueness" that can give a person some important wiggle room if they are not able to prepare for all contingencies.

not a document meant to be viewed as a part of a person's past, but rather as if it was the patient speaking the words with her own voice at the time questions arise.

In this chapter I will argue that the extended mind makes it the case that the directive should be interpreted as little as possible. We should take a directive at its word. We disrespect the autonomy of the drafter the more we stray from the words on the paper. The words on the paper are an isomorphic copy of the contents of a person's biological brain – as far as we can possibly tell. No level of neural or linguistic analysis can yield a better understanding of the wishes of the patient than the patient is already able to give. The directive is written with language, understood with language and any attempt to reexamine the directive will have to be done so with language.

That being said, there are nonetheless times when a serious departure from the text is not only necessary but moral incumbent. These would be cases when there is an appropriately placed proxy who is in the best possible position to reveal what the patient would desire. This best of all possible positions comes in the form of what Clark and Chalmers (1998), Rovane (1998) and Nelson (2003) call 'unusually interdependent couples'. These are cases (perhaps rare cases) where the patient and the proxy participate in the same mental life in some regards. Each retains an individual identity but, in some way, share the mental resources of the other for specific purposes. In these instances, the proxy has complete authority over the interpretation of the directive, even if it means a drastic departure from the express wishes in the directive. These kinds of relationships may be rare but, even if they were quite common, there would still be a need for proxies which do not rise to quite this level of interconnectedness. Determining who is the right, or at least the most appropriate proxy will be concern addressed later in this chapter.

Three steps will be necessary to carry out the goals of this chapter. First, I will discuss the isomorphic character of some parts of language. That is to say, language on paper is functionally the same as language in the brain for purposes of interpreting what the speaker means. If there exists vagueness in a written directive there is good reason to suspect the same vagueness existed in the mind of the drafter. Secondly, I will apply the four criteria for extended cognition as a means of determining the suitability of a potential proxy. Proxies are necessary in both cases where an advance directive is in need of interpretation but also when a directive is absent as well. The same criteria are used in both cases and is applied the same way. Lastly, I apply these criteria again, this time to the advance directive itself to further demonstrate the directive as an externalized cognitive function of the patient. For now, we need to look at the medium through which advance directives are made known and medium the debate over their implementation occurs: language.

Isomorphism of Language and its Limits

> Language is the liquid
> that we're all dissolved in.
> Great for solving problems,
> after it creates a problem.
>
> Modest Mouse
> *Blame it on the Tetons*

What follows here is the argument promised in the previous chapter regarding the isomorphic nature of language. To begin we will first look back to what the first thought experiment by Clark and Chalmers in *The Extended Mind* has to do with a video game remarkably like Tetris (so remarkable that I will refer to the game as such).[118]

[118] Just in case you are unfamiliar with Tetris here is an image to help demonstrate the task at hand:

In this experiment we are asked to imagine three (slightly) different tasks:

1) A person mentally rotating a Tetris block on the screen to determine how many 90º turns clockwise it would take to make the shape fit in a given hole.

2) A person determining the number of turns, but with the aid of a button which, when pressed, will rotate the shape 90º clockwise on the screen.

3) A person with a neural prosthetic which allows them to turn the shape on the screen by merely thinking about doing so, rather than by physically pushing the button.[119]

The conclusion the authors draw from this thought experiment is that if we consider case numbers one and three to be a mental process we have no reason to exclude case two from the same group. The only difference between case one and case three is the presence of the neural prosthetic and the only difference between case two and case three is the location of "the

Figure 4- Illustration of possible moves in a game of Tetris

The black shape represents the image to be rotated, the dark gray space represents the existing blocks and the light gray represents the space to be filled by the incoming black block.

[119] The technology to do this exists. I do not know if the technology has been used for exactly this purpose, but see (Hochberg, et al, 2012) and (Serruya, et al, 2002) for other examples.

Additionally, Stelarc, a Cypriote born, Australian performance artist, has been working with human machine interfaces since the 1980s. Stelarc uses electrodes places on the exterior of his body (often on the abdomen) and learns to control impulses sent to different muscles to exercise control of his robot creations. Though technologically different from the other cited examples, his work is nonetheless an interesting and enlightening demonstration of the possible connections between humans and machines. (see Clark, 2003, MotherboardTV, 2016 and TEDxTalks, 2016)

button". Ergo the button in case two should be, and is, part of the person's mind, therefore, the mind extends outside of the brain.

I do not believe this thought experiment is itself a slam dunk for the extended mind thesis (though it is compelling). The thought experiment is not without its problems (which the remainder of Clark and Chalmers' paper addresses, and I do not believe need to be discussed again in this dissertation). I mention the Tetris block example so that I may take it a step further. Whereas Clark and Chalmers are concerned with the physical means by which the subject manipulates the world in terms of the location and operation of the tools used, I would like to focus on the contents and form of the mental activity that initiated the change.

In all three cases the test subject desired to determine the number of times the block had to be rotated to fit into the given space. What differs is the means by which they fulfilled this desire. In case one the desire was fulfilled without making any changes to the external (relative to her skull) world. The subject mentally (that is to say, in her biological brain) rotated the block and internally kept a running count of the turns. Here the desire led to (and only to) a new mental state; no change in the world apart from her internal brain state has changed. In case two the desire was fulfilled when the button was pressed by hand and the block rotated 90º until the correct orientation was achieved. Here the desire led to both a new mental state and two physical changes (the hand moved, and the block rotated on the screen) in the external world. And finally, in case three the desire led to a new mental state and one change in the physical world (the block rotated but in this case the hand was not involved).

What is internal and external (to the subject) in the performance of the task is different in all three cases. Some of the mental processes and results were externalized in some instances and not in others. These cases all appear to be different ways of achieving the same result. The

means are different enough that we might rightfully not considered them all the same kind of action. In some cases, the actions were purely mental (i.e. – happening wholly inside the head of the individual), while in others not. Though the result might be a hole in the ground, there is a world of important difference between using one's hands, a shovel, a backhoe or dynamite.

What these cases have in common together is the sameness of the desire that led to the result. In all cases a mental activity was initiated by the subject. Let's consider a different, purely internal solution to the question of how many turns are needed by looking at three approaches based on the image from footnote #133. Subject one mentally turns the block clockwise 90º one time and sees that it fits. Subject two makes the mental rotations counter clockwise and gets a result of three turns. A third subject considers both possibilities and concludes that the answer is one or three depending on which direction you choose to move the block.[120] The mental route taken by all three subjects is slightly different, but we would not say that any of them had engaged in a fundamentally different activity than the others.

Another example is the difference between how my sister and I learned (and currently do) basic addition. She tells me that she does this by sheer force of memory. Like multiplication tables, she just memorized the results of the addition table. I, on the other hand, developed a heuristic for at least some cases. If a single digit number is added to 10 the result it the same as placing a 1 in front of the number (10+5= 15, which is similar to the standard practice of moving the decimal point in base ten counting systems). Obviously, this heuristic does not work when

[120] In most versions of Tetris, the player has the option to spin the block clockwise or counter clockwise by using different buttons. There is a practical implication here in that at high speeds three turns one way might take too long compared to a single turn in the other direction. This recognition of which is the faster direction of rotation is a major factor in separating the novice players from the elite. For the purposes of this thought experiment, however, time is not of the essence.

adding 9s, such as 9+5. Instead of memorizing the results I subtracted 1 from the addend (5 in this case) and follow the procedure used for adding 10 (e.g. 9+5 becomes instead 10+4). It's a little convoluted, but it works for me. Are my sister and I doing something fundamentally different when doing addition? Perhaps, she is engaged in an act of memory recall, whereas, I am doing a calculation which uses subtraction as well as to addition and, therefore, is somewhat more complex than necessary, but in the end, from an outside perspective, there appears to be no difference.

Currently we do not have a full grasp on how states of the matter that makes up the brain result in consciousness, memories, desires or actions, but the fact remains that they do (unless you are going to hold to some kind of substance dualism; a theory which I do not believe needs further debunking from me). If I draft an advance directive I do so using my knowledge, memories and skills, all of which are "found" (initially) in my brain. I have a desire to draft an advance directive which says X, Y and Z. When I put pen to paper this desire is now 'written' in two places, in my brain (via encoding which produces an engram, under Tulving's model as discussed in chapter 2) and on the paper. I desired to write and the desire can be assumed because something was written. If we accept what is in my brain, why not what is on paper? They are the result of the same desire so if there is a difference, it would be found in the information itself. The question is not whether or not the words on the paper are my words but rather to what extent the words on the paper are isomorphic with the thoughts in my head. We accept my wishes as I vocalize them at the needed time. All things being equal, words on paper are the same words which I would utter, were I able to at the needed time.

One objection might be that, despite the claim just made, there is a world of difference between what I think and what I write. Language is a limited medium. It is no small feat for an

author to capture the phenomenal essence of an experience, memory or sensation in mere words. We do lose something when we put thoughts to paper versus thoughts to words. Tone of voice is difficult to capture in written language, but it is part of the basic mechanics of the spoken word. Isn't there more to my thoughts about end of life decision than what I can spill out onto a piece of paper? Yes and no. Language may be a *poor* medium in some cases but, alas, it is the *only* medium we have with which we can describe our thoughts or make our thoughts known to someone who does not have the same privileged access to them while they reside solely in our skulls.

Say that, in the future, we have the ability to take a snapshot of a person's brain and with that snapshot we can translate the neural connections, chemical levels, intensity of impulses, or any other relevant physical state of the brain into a narrative which describes exactly what was going on in that person's head as the person understands it. We philosophers talk of qualia, the what-its-likeness to have a certain thought or mental state. When I taste strawberries, I feel a certain way. I do not experience this as "my neural connections are such that I am tasting and enjoying strawberries". There is instead, just something it is like for me to enjoy a strawberry which results from the physical states of my brain. How things get from this complex set of physical states in my brain to my conscious experience is not clear (so unclear in fact that it is called one of the 'hard problems' of consciousness). For purposes of this snapshot we do not need to know how the translation from brain state to quale happens, just that brain state X results in the experience of qualia Y and that a brain state is something we can, in principle, measure and map. With enough information about a brain's physical state we could (again, in principle) deduce the mental state (both conscious and unconscious) of the person. Let's say a snapshot of my brain is taken right after I drafted my directive. Shortly after the snapshot is taken I lapse

into a PVS. What would we find if we compared the contents of the directive with the contents of the memory derived from the snapshot? Surely there would be some differences. Time has elapsed, so I would have a slightly different set of memories. My current sensory experience would be different since I was looking around the room rather than at a piece of paper. In such a short period of time it is unlikely I've completely changed my mind about what I just wrote in my directive.

When we examine the snapshot, we find neural connections and chemical states which translated, to me, as the instructions I wanted to convey in the advance directive. The reader of the snapshot would not experience these same physical states in the same way I did but nonetheless would be able to describe what I was experiencing. This description would be carried out in language, with words. The words on the paper which constitute my advance directive are "found" in my head in the form of some brain state. My brain and the directive contain the same information, albeit in a very different form. It is a complex translation for the interpreter of the snapshot, yet almost effortless for me. I can no more make a listener feel the what-its-likeness of drafting an advance directive by speaking aloud than I can by writing the same instructions. The information qua information is isomorphic across neural and external media.

What reason do we have to believe that an excruciatingly close examination of my brain will yield different facts about my mental state than an examination of my competent reporting on those mental states? Yes, my first-person experience of these mental states are not captured by the words on paper but no amount of description by me, no matter the means of transmission, can ever fully get across the qualia as I experience it. If this is the case, then all things being equal, if we do not second guess what a person says so long as they are competent we should not

second guess what that person writes. The same information can be equally discerned between the medium of speech, written language and (assuming our super science future) snapshots of brain states. Beyond that, it is not obvious that what is missing is always relevant or required when considering one's words.

If we cannot trust an advance directive as being a record of what a patient desires we, by the same logic, will have a hard time accepting those same wishes if uttered in the moment, as they are both externalized reports of brain states. Speech is far more ephemeral (it disappears from the non-mental world almost as soon as it is uttered) than a written document. But there is still a record of the speech nonetheless. My orally expressed desires are recorded in the brains of those listening. My written desires are first placed on paper and later (maybe a great deal of time later) are recorded on the brain of the person who reads the document. If my wishes and memories are what matter in cases of moral decision making it should not matter how those wishes are expressed, as the result in all cases is ultimately to end up as a mental state of the person in receipt of the information I wish to convey.

Conclusions

In the 1999 film *The Matrix* the character, Cypher, spends much of his time examining green code cascading down a black background. Though seeming nonsensical symbols to most, Cypher has learned to 'read' the code saying:

"You get used to it. I, I don't even see the code. All I see is blonde, brunette, redhead. Hey, you eh, want a drink?" (Matrix, 1999)

Cypher has, matter-of-factly, summed up the dilemma of translation of thought into any external media. While he sees squiggles and squaggles of the matrix language on the screen and internally experiences these as various qualia, he can only explain the sensation in language; 'this line of code is a redhead'.

Whether or not humans "think with words" is a controversial issue, but what is not controversial is that fact that we must (in the sense that no tool is available to us to do otherwise) use words to talk about thinking and thoughts. Regardless of the medium through which a person makes the contents of her mind known, she will have to do so with language in one form or another. A possible exception being an ability to look directly into her brain and translate those physical states but even then we have to interpret, discuss and debate those contents with language.

While our need for language to convey both simple and complex ideas creates a limitation on exactly what information we can disseminate, it also makes it the case that information can retain fidelity across media. Though tone of voice plays an important role in verbal communication[121] in cases where tone of voice is not relevant or appropriate (advance directives, I believe, fit this bill), information can be said to perfectly survive translation across media. This being so, an advance directive is not a second-best record of what is happening in a person's mind, it is what is happening in a person's mind, though the location of the thought is not what we traditionally think of as being where "the mental" if located.

All of this gives us good reason to take an advance directive 'at its word'. If a person is the best source for what that person's desires and wishes are, we have no principled reason to favor an oral over a written request (or vis versa). The patient's desires are the patient's desires regardless of the medium in which we encounter them.

This may mean that a patient 'means what she says' but we are still, sometimes, left to decide what exactly was said in the first place. There is an escapable vagueness in language and

[121] Consider the sentence "what are you doing here?". The meaning of the sentence changes depending on what would is emphasized. If no emphasis is used the information contained is isomorphic and perfectly survives translation across media.

as such, medical personnel will, from time to time at least, need to enlist the aid of an outside party in determining what *exactly* the patient was driving at or how would the patient respond to her situation given new information. So, we now turn to the question: how do we choose a proxy?

The Extended Mind and Choosing a Proxy

In cases where a patient is unable to make medical decisions for herself it is necessary for a proxy to make them for her. An ideal proxy would be one appointed by the patient prior to her loss of competence and who would choose either as the patient would choose or always act in the best interest of the patient.[122] Obviously, the strength of relationships between people varies greatly. A court appointed guardian who has never met the patient will be in a far weaker epistemic position concerning what a patient would wish given her circumstances, while a lifelong partner would be in a relatively superior position in this regard.

It is not always possible to appoint a proxy ideal suited to the patient's needs. Whether there is or is not an obvious choice for a proxy the extended mind thesis helps in determining what persons and choices should be privileged over others. In this section I consider cases where one or more possible proxies are available.[123] The extended mind thesis gives guidance and justification for picking one proxy over another as well as added strength to determinations of that proxy. In fact, in light of the extended mind thesis, in some cases a proxy is not a proxy per se. If the extended mind theory is true, the terms 'proxy' and 'proxy decision maker' are

[122] What is in the best interest of the patient and what the patient would choose are not always the same thing and there is some debate over which of these is the proper role of a proxy. As this section goes forward I will argue that the best of all possible proxies would choose as the patient would choose and thereby, act in (what the patient sees as) her best interest.

[123] Cases where no obvious proxy is found, or the appointed proxy is a stranger to the patient, are the topic of the appendix at the end of this dissertation.

sometimes misnomers. If it is the case that a single mind can span two (or more) brains, one person is not serving as the proxy for the other but is rather a second medium by which the desires of the patient can come to be known. If this spanning is possible, then we are very close to resolution in the debate over the legitimacy of proxy decision making in some cases.

Consider, yet again, the phone number example, but with a few variations. I do not need to remember a phone number, rather the person to my left (Lucas) does. I do not believe I will be able to remember the entire phone number, so I enlist the person on my right (Roland) to help. Lucas gives me the 7 digit number and I relay the 4 digit line number to Roland but fail to tell him what the numbers are for (could be a phone number, address, lottery picks, PIN, etc.), but I do tell him that Lucas may ask him for these numbers someday and that it is my wish that he pass them along. Lucas later asks me for the phone number. After giving him the 3-digit extension, I pass out. Later, upon request, Roland gives Lucas the missing information.

It was my intention to be able to give Lucas all of the information he required at the appropriate time, but I was unable to do so using the body and brain we orthodoxly consider 'me'. Knowing it was a possibility I might become unable to carry out my wish to give the numbers to Lucas, I made a contingency plan for my wishes to be carried out. Roland is the means by which this would happen should I be unable. Roland is not making a best guess at what the information I would give would be, but rather revealing it in the same manner I would. For the purposes of remembering and relaying the phone number I have incorporated Roland into my cognitive machinery. Not only is Roland revealing the numbers, but so am I. Roland is carrying out a linguistic task which is purely isomorphic. Regardless of the reasons, Lucas needs the numbers and Roland can report them with complete fidelity. The desire for Lucas to have this information is mine and remains mine. Roland may also have a desire to see Lucas receive

this information but given the situation as it began and how the information disseminated through the parties involved, it was my desire to help Lucas which led to Roland's involvement in the first place. Without my desire the information would not have been passed along. Roland is capable of completing my task in my absence even if he did not know what the information was about in the first place.

The reader will rightfully question the aptness of this analogy. A phone number, a mere 7 digits, is as isomorphic as information gets. It is a leap to go from the claim that relaying simple information such as a phone number is a cognitive task that can span brains without loss of meaning to the claim that an advance directive can be interpreted by someone other than the drafter without loss of meaning, intent or force.

There is precedent in biology for considering not only the properties and functions of an organism itself but also within the context of social position. Just as we cannot fully understand ants, termites or flocks of birds solely by examining a token of the species, neither can we fully understand 'persons' without looking at token persons in their social context. In what follows I will discuss the idea of group minds and group cognition by considering analogies with groups of social, non-human animals. Clearly, the interaction between members of the human species are more varied and complex than those seen in social insects, but the fact remains that our social nature opens the door to plausible levels of analysis of what constitutes our mental lives beyond what is occurring solely in our heads.

Group Minds and Group Cognition?

I have repeatedly called upon the "phone number example" in the course of this dissertation. By my own admission it is a trivial example but, nonetheless, I think illustratively useful. I will now need to make the jump from this trivial adoption of another's mind to the

more complicated and less obvious claim of a person spanning two bodies.

Clark and Chalmers discuss this in *The Extended Mind*, but only as a possibility:

> What about socially extended cognition? Could my mental states be partly constituted by the states of other thinkers? We see no reason why not, in principle. In an unusually interdependent couple, it is entirely possible that one partner's beliefs will play the same sort of role for the other as the notebook plays for Otto. What is central is a high degree of trust, reliance, and accessibility. In other social relationships these criteria may not be so clearly fulfilled, but they might nevertheless be fulfilled in specific domains. For example, the waiter at my favorite restaurant might act as a repository of my beliefs about my favorite meals (this might even be construed as a case of extended desire). In other cases, one's beliefs might be embodied in one's secretary, one's accountant, or one's collaborator. (Clark and Chalmers, 1998)

If my waiter was actually sharing in the same mind as me this would be customer service taken to a totally new level. However, I will focus on the "unusually interdependent couple" later in this chapter rather than these kinds of specialized relationships, as it will be the most fruitful and relevant relationship for the topic at hand.

It is not only in humans we see a discussion of the possibility of group minds or mass organisms. In the contemporary biology literature there exists a discussion of the possibility of a group or 'hive' mind among certain social insects, such as bees and ants. This is germane to the present topic in at least two ways. One, it opens the door to the possibility that organisms (including humans) might not be the discreet, individuated entities as we tend to think of them. And secondly, failing this, the door is now open to a different level of analysis of organisms. We can learn a great deal about a species by studying one of its members, but such a study cannot tell us everything. Part of what an ant is in part its relationship between other members of the hive. The same is true of humans or persons. Part of what makes a human a human is how it is situated within its society, culture, family unit or social group.

This kind of thinking reaches at least as back as far as 1910 when William Morton Wheeler suggested parallels between the organism as organism and colony as organism

(Wheeler, 1910). E.O. Wilson and Bert Hölldobler discuss this issue in detail in 2009's The Superorganism: The Beauty, Elegance, and Strangeness of Insect Societies. In it the authors espouse the view that colonies of social insects (ants, bees, termites, etc.) themselves constitute organisms above and beyond the constituent critters that make it up, even going so far as to draw parallels between that kind of ant (and ant activity) and parts of more orthodox organisms (i.e.- cell=colony members, organs=castes, gonads=reproductive castes, etc.).

Ants are part of a functional system which defines what kind of thing each individual is. I wish to make it clear that Wilson and Hölldobler are not making the argument that a token ant is not itself an organism, nor that the colony is a life form independent of the insect which constitute it. An ant itself is just as much an organism as a human being, squid or a ficus. They are making the case that there are at least two different levels of analysis when talking about social insects, the individual and the group.

Humans are social animals. We have the longest period of dependency on older conspecifics of any animal (regardless of average lifespan) and prolonged isolation from social contact is often psychologically harmful. Though we are many orders of magnitude more complex than ants or bees we are nonetheless defined, in part, by our relationships with other members of our species. If we can examine social insects at a group level to learn something about the individuals, perhaps we can do the same thing with human beings. There already exists a constellation of levels of analysis by which we try to understand what a person is. John Locke views a person as a thing with a continued memory, Thomas Reid is in agreement with Locke, but adds a biological sameness which he thinks is logically required. David Hume argues that 'person' is a matter of convention. More recently, Carol Rovane (1998) talks of 'person' as denoting an ethical kind in addition to a biological or psychological kind.

It is appropriate to consider the meaning of 'person' differently in different context. If in fact group minds are possible among humans, there is still reason for viewing each human component as biologically distinct for purposes of treating disease. It would not serve a person with cancer well to give chemotherapy drugs to a cancer-free person with whom she shares an apartment. However, the landlord could collectively think of the two together as "the tenant" and not care who pays how much of the rent so long as the rent is paid in full each month. I believe any argument concerning the full integration of two token individuals into one person in all meanings of the word 'person' is suspect at best and a fool's errand at the worst if for no other reason than the animal identity of the tokens remains distinct. As we discuss group minds in this dissertation I will focus on those meanings of the word person which still allow for distinction of individuals when the context necessitates.

I would like to borrow a case purposed by Nelson (2003) regarding a Jehovah's Witness and an emergency blood transfusion. Nelson refers to this patient as Martha and I will follow suit.[124] Martha, an elderly woman, has been a devote Jehovah's Witness her entire adult life. Her devotion is so absolute that she accepts the word and teaching of the church unquestioningly. Over the years as church dogma has changed so too have Martha's personal views. She falls in line with church teaching, in lock step, no matter the position. When she learns of a change in doctrine she consciously changes her position to match.

The Jehovah's Witnesses have an infamous belief that blood transfusions are an unacceptable medical practice, referring to it as 'blood eating.' (Evans, et al, 1973) Martha, in

[124] In Nelson's version of the case Martha begins as a "moderately demented woman who is currently living a reasonably pleasant life." Dementia is not an important component of my use of the example. We will assume before illness she was a perfectly competent individual.

need of routine, yet necessary surgery, makes her belief on this issue know to medical staff by issuing a written directive. If she is to need a blood transfusion during the course of her treatment it is not to be given to her as this would be counter to the teaching of her faith and thus her conscience.

During the procedure, there are, of course, unforeseen complications during her surgery and blood is needed. In most cases this would be a clear-cut case of an advance directive given by a competent patient who is currently unable to consent. Medical staff refers to the last known competent instruction and carry it out. However, when the family is informed of the situation they tell the doctor that while Martha has been unconscious the leaders of the Jehovah's Witness church have officially changed the position of the church on this issue. Blood transfusions are now permitted when it is necessary to save the life of the patient. The family instructs the doctor to go ahead with the procedure because this is what Martha would now want.

Have Martha's wishes to not receive treatment been denied or has Martha changed her mind about what amounts to acceptable treatment? In light of the new doctrine of the church has Martha changed her belief even though she is not aware of it? In a Facebook conversation I received two comments that represent a wide swath of opinion counter to the idea that Martha's beliefs have changed:

Michael Zois: In the hospital, no. Unless she has left an advanced directive saying she will accept changes in church dogma, her last good instructions are binding.

[Chuck] Watt: But the position isn't changed until accepted, which is still an act of free will, no? The acceptance is what creates compliance. If the JWitness (sic) totally went of (sic) the rails and switched to a "Heaven's Gate" type death cult during her coma, would we infer her pledge of mutual suicide?

Michael's objection is pragmatic (no doubt drawn from his time working in hospital administration and as a surgical first assist). We have a standing directive which is clear in what

the patient wishes to be done, or not done, in this particular case. From a legal standpoint I do not think anyone would begrudge the hospital for carrying out Martha's directive, should no one close to her object. It would be an onerous fiasco to require hospitals to review each patient's directive to ensure agreement between it and whatever up to the minute world event might have played a role in the patient changing her mind. This is in part because the hospital is not a good proxy decision maker for the patient (by failing to meet the extended mind criteria to be discussed shortly). Though hospitals and the health care workers therein have an interest in seeing patients get what is best for them, hospitals are not good at determining what the patient *believes* is best for them.

In Martha's case we have an appropriate placed proxy at the ready to better interpret her wishes. The family members here are not choosing for Martha contrary to her wishes. Instead they are revealing what her wishes would be were she able to take advantage of this new fact (the change in church doctrine). Given Martha's history, the family knows that there are at least two sources for inquiring what it is Martha believes on some issues. The first is, Martha herself. If you want to know what Martha believes, Martha is the best of all possible sources, so long as Martha is competent. But given Martha's past endorsement of the teaching of the Jehovah's Witness church we can also make an argument that these documents and oral teachings constitute a second source of information about Martha's beliefs. Since Martha has, in the past, accepted so completely the teachings of the church we have no reason to see them as different sets of beliefs but rather one set of beliefs carried or stored in two different places. I have a less controversial analogy which may help the reader understand the point.

I believe the algorithm of multiplying numbers of two or more digits to always yield accurate results when applied correctly. There will, in all likelihood, be a time in the future

where I will need to multiply such numbers which I have never multiplied before; for instance: 123 X 24.

If I ask someone to calculate this product for me and trust they are applying the proper algorithm with sufficient skill, I will believe the result even if I have not done the calculation myself. Even if this information went through several intermediaries I would nonetheless endorse the result so long as I had reason to trust those calculating it and passing the information along. My endorsement of the result is not dependent on my personal calculation of the product but is implicit in my past endorsement of the algorithm and trust of those doing the calculation. So long as I endorse the algorithm I endorse the results it generates without having to ever endorse any particular result or other.

If I were in need of the product of 123 X 24 and was rendered unconscious before being given the result, anyone close to me who knew of my endorsement of the multiplication algorithm would be correct in saying "Michael believes that result" so long as the person deciding this for me likewise trusted the skill of the mathematician. I do not recall ever explicitly saying "I endorse the standard multiplication algorithm,"[125] yet I have tacitly endorsed it perhaps thousands of time in my life by using it and applying the result. Likewise, Martha may never have said "I always change my beliefs to match those of the church" yet every time the situation arose she has done so. It is not the least bit unreasonable to induce that Martha's future choice in this area would conform with her past behavior.

Chuck's objection concerns whether or not Martha would accept absolutely any change made by the Jehovah's Witness church no matter how divergent it is from teaching in the past. It is not impossible to see a scenario in which Martha does come to embrace the change from a Christian

[125] At least not until writing this sentence.

faith to a death cult, so long as these changes take place in rational stages, perhaps over time. After all, the Christian church today is far different than the Christian church 100, 200 or 1,000 years ago. Even within the span of a single lifetime a worldwide faith can undergo significant changes (e.g. – The Second Vatican Council). The endorsement of a person, idea, algorithm, or what have you, need not be absolute and never wavering. An endorsement does not mean one can never recant or have a difference of opinion in the future.

In Martha's specific case we are not seeing a radical change in church doctrine. A repeal of the prohibition against blood transfusions would be an adjustment to a single rule, not a fundamental change in the structure of the organization. Arguments in other areas can be made that we often believe things without explicit endorsement if there is a chain of logical consequences which lead to the belief.

Though a contentious view in the astronomical community, the demotion of Pluto from 'planet' to 'dwarf planet' serves as good analogy. As hard as this switch in astronomical "belief" was for some, it was not nearly as difficult as it would have been should the International Astronomical Union (IAU) declared that only rocky bodies are planets or that all and only planets support life. If I now believe Pluto is not a planet there are a number of other belief which I now have which I may well never explicitly endorse.[126] For example, there are eight planets in the solar system, that Neptune is now the most distant known planet from the sun and that the mnemonic "My Very Excellent Mother Just Served Us Nine Pizzas' no longer makes sense as it once did. What this reclassification of Pluto does not imply is that I now hold a

[126] Of course, I have now written them down therefore, have considered them and have hence accepted them as true. I respectfully ask the readers to put this little bit of temporal logic out of their minds.

geocentric view of the universe that the moon is made of green cheese. These beliefs bear no logical relation to the altered belief about Pluto's status.

A Note about Death Cults

Chuck Watt's comment concerning the possibility of the Jehovah's Witness church turning into a death cult overnight is not an easy objection to dismiss, as unlikely as the possibility is. I will try but, in the end, there will be some bullet biting.

Even not being an expert on the Jehovah's Witnesses, I believe I can still show a logical chain of belief that concludes with a change in the blood transfusion doctrine. It follows thusly: though blood transfusion and organ donations have been treated as cannibalism by the church in the past, these medical procedures are sometimes necessary to sustain human life. The protection of human life is also an important virtue and not allowing these medical procedures could end rather than preserve a life. The new conclusion is that so long as the blood and organs were voluntarily given and did not result or cause the death of the donor the act is something different than cannibalism and is thus permitted. I cannot, for the life of me, find any argument which moves the Jehovah's Witnesses to a pact of mutual suicide without some torturous logic and massive assumptions of what Jehovah desires, therefore this move apparently fails this test involving logical relations.

However, Martha did not change her mind regarding blood transfusions based on the logic of the church's argument or logical examination of the move after the fact. We say Martha changed her mind because she has a strong history of doing so without such reflection. Her reasoning is merely, 'this is right and I believe it now because the church says so.' This would be as if I changed my mind about Pluto's status only justifying my position by saying "because the IAU said so." If this really is the only justification that Martha has given she may well

believe that the Jehovah's Witnesses are now a death cult. Her reasoning is not sound by any reasonable understanding of the term, but it is nonetheless the reasoning she has used all her life.[127] Should her advance directive instruct us to refer to the latest proclamations of the church and following them in making decisions for her we have even more reason to think she has changed her mind thusly.

Bullet bitten.

Choosing a Proxy

Just because a belief never finds its way into conscious memory or is never explicitly endorsed does not mean it is not believed. The previously mentioned astronomical beliefs are a consequence of my new belief that Pluto is not a planet, not of my acceptance of each individual belief. If Martha, through explicit declaration or tacit endorsement, gives reason to believe her beliefs are always those of the Jehovah's Witness church we have good (maybe impeccable) reason to say that Martha now believes in the permissibility of blood transfusion, so long as the reasoning behind the change is relevantly similar to changes she has endorsed in the past. The belief in the rightness of blood transfusion is a consequence of Martha's belief in the teachings and doctrines of the church.

Again, in Martha's case we have people close to here giving an interpretation of her advance directive. Though they may say "in light of the changes purposed by the church Martha would now accept a blood transfusion" they can just as plausibly say "because Martha believes so strongly in the teaching of the church her opinion of the matter has changed, though she is not aware of the change." We would not fault the caregivers for following the directive in the

[127] Exactly what the implications are for Martha in her current situation is unclear, but I doubt it's very pleasant.

absence of a third party able to give a new, reasonable interpretation as the mentality of the care givers is not connected to Martha in the way the mentality of an appropriately placed proxy is connected.

The proxy has not changed Martha's mind, but rather *revealed* that Martha had already changed her mind. We have every reason to believe Martha would have made exactly the same choice given the same input. What is it that makes an appropriate proxy? The obvious answer is the "closer" to the patient the better. But this, itself, will not be sufficient, in part because "closeness" can mean many different things. How do we resolve issues of two persons with equal claims of closeness with differing views on how the patient would decide were she able? What about cases where there is no one significantly close to the patient? What if the purposed proxy is close to the patient but has an obvious agenda which is not aligned with the patient's interests? Since we are here looking at the implications of a theory of mind we could do worse than to begin with examining that theory for answers to these questions.

To reiterate, according to Clark and Chalmers and Levy in order for an external apparatus to be part of a person's mind four criteria must be met. The entity:

- is readily available
- is easily accessible (transparent in use)
- is a constant in one's life
- must have been endorsed in the past and there is some consequence to its future endorsement. (Clark and Chalmers, 1998 and Levy, 2007)

People are not notebooks or wristwatches, but nonetheless I think these criteria will serve well in determining the viability of a proxy. Some of the criteria might have to be differently understood, but not torturously so.

I will now discuss how these criteria apply to proxy decision makers as well as make an attempt to determine when the criteria have been met. As with their more traditional application, none of these criteria are alone sufficient for establishing an extended mind but they are all, to one degree or another, necessary. Taken as a whole I think a compelling case can be made that the best proxies participate in a (at least partially) shared mental life and that such a shared mental life is possible.

Readily Available

A proxy who is present is best, but this is not required. In the past presence was vital for those making decisions for others. Without so much as a phone (or perhaps a telegraph, if I could extend this idea further) someone on the other end of the country could never be aware or a loved one's acute condition let alone be able to make decisions for that person in any reasonable or useful amount of time.

Clark has much to say about presence without physical presence. In his books Being There and Natural Born Cyborg he discusses the possibility that our physical presence may not be necessary in a number of circumstances. I have had a small handful of meetings with committee members while writing this dissertation over Skype and I do not believe the introduction of this technology, versus face-to-face contact, had any appreciable effect on the quality or productiveness of those meetings for good or for ill. But there are limits to the power of telepresence. If I have never physically been to Cambodia but viewed the country through a live feed via a camera attached to my friend Eric's glasses there seems to be something amiss if I claim to have "been to Cambodia." Even if we created an interactive virtual environment complete with a 360° view and oppressive humidity, there is still something untrue about my claiming to have been to Cambodia. We do not feel the same connection to an event watching it

on TV was we do being there to witness it (like Woodstock, the World Series or Tiananmen Square). I have seen dozens of videos of Nirvana in concert, some of which were filmed in places I have been since their appearance there, yet I feel a certain sense of emptiness at never having been physically present at those places while the band was present.[128]

These controversies aside, there are plenty of times we endorse the actions and respect the wishes of persons not in our immediate company. We, under most ordinary circumstances, accept what a person says over the phone as her words just as we would were she standing right in front of us. A faxed copy of a signature is, for many purposes, just as good as signing in the presence of the would-be fax recipient. This leniency only goes so far, however. In legal cases taking a person's "word for it" is sometimes not enough or a notary may be needed to verify a signature's authenticity.

In the end, distance between patient and proxy is not the important issue here. What is important is the presence of the proxy *in the person's life*. As it stands now, in most jurisdictions, next of kin, regardless of the distance of the relationship, is the default for appointing a proxy. This might be legally useful but is also morally dubious. An estranged uncle is a poor choice when compared with a life-long, unmarried, romantic partner or close friend. The closer relationship should be preferred, as we are more likely to arrive at an interpretation most in line with the wishes of the patient.

The time in which the proxy need be readily available is not just at the time of decision making. The proxy needed to be available to the patient throughout her life so as to create the cognitive bonds that make the extension of the mind over separate brains possible. The proxy

[128] My dissertation advisor made a comment here: "Alas, you could never really say 'Here we are now. Entertain us'." This was, no doubt, meant to be a throwaway gag to get a chuckle out of me but which nonetheless illustrates the importance of presence over virtual presence.

needs to have been present as the patient made hard choices in the past, perhaps even participating in the process. She needs to be aware of how the patient would act by having witnessed to the patient's behavior in the past. The presence of the proxy should have been fluid, organic and easy rather than contrived or actively pursued, despite the potential proxy's protestations. The extent to which the proxy sees the patient's problems as her (the proxy's) own personal problems is also a good indicator.

Easily Accessible (Transparent in Use)

As with the previous criteria, distance is not really the issue here. This criterion applies more as a means of evaluating the appropriateness of the proxy's decision as much as how easy it is to make that decision known.

Obviously, a conscious, rational person makes a better proxy than an unconscious, incompetent person. A conscious and rational person meets the criteria of being easily accessible in the same way the assistant from the phone number example does. If you want the answer, just ask them. This, however, will not be enough to fill the criteria "transparent in use."

We need to apply the idea of 'transparent in use' to the quality of the determination made by the proxy. This is a determination of appropriateness after the fact. If we believe a person is a viable candidate for the position of proxy but the decisions they make are illogical, unsound, unclear or hedged we have reason to rethink the person's qualifications. The patient, while she was able, was the best judge of her own mind (such is the nature of first person access). But if the proxy's brain is serving as scaffolding for the patient's mind we can assume a (much) better than average understanding of the patient's desires or wishes than a potential proxy whose brain is not being scaffolded.

This is not to say that the decision made by the proxy needs to be made without hesitation or moral qualms. Certitude from the proxy may be telling, but it is not everything. Despite arguments that competent people always act always with their best interest (or at least what they believe is their best interest) in mind, they do not always do so without deliberation or doubt. For example, should I undergo chemotherapy to prolong my life despite the likelihood it will be prolonged in a more painful or uncomfortable state? This is not a choice everyone could make quickly and with certainty. Some self-reflection when making big decisions is to be expected and is not a sign of ignorance of one's self interest.

A better way to view 'transparent in use' in this case would be to consider ready availability of the interpretation of the directive in the mind of the proxy. As with self-reflection, some degree of soul searching, and doubt, is expected when making difficult medical and moral choices, but a closeness of relationship might be assessed by how readily the proxy understands the relevant points and the conflicts the patient would recognize as important. If we have a choice between a unmarried partner and an estranged uncle debating the merits of terminating life sustaining treatment, we can look at their reasoning to determine who most likely understands the desire of the incompetent patient. Say, for instance, the uncle turns to scripture and asks that the treatment continue, whereas the partner knows the patient has never turned to scripture to make any decision in her life. The uncle demonstrates a cluelessness as to the desire of the patient while the partner shows some understanding, thus making her decision more readily available to the patient.

But this begs the question. Given the patient is unconscious or otherwise unable to say for herself what it is she would desire, how are we to know which of these potential proxies is closer to participating in the same mental life of the patient? If transparency was the only criteria

to consider this may well be an insurmountable challenge. However, none of these four criteria are themselves sufficient for determining the presence of an extended mind, though each is necessary. The other three criteria are easier to evaluate objectively. Readily available (as tweaked for our purposes) can be determined based on whether or not the potential proxy can be located and contacted. The constancy in the patient's life can be determined comparing the longevity and immediacy of the relationship between the patient and the potential proxies. Lastly, endorsement of past consequences likewise might be induced as evidence may exist of past endorsement.

To this point I have been speaking of the proxy's decision as being qualitatively identical to that of the patient, were she able to make it known. In truth this is just a shorthand for sameness of cognitive style or use of given resources like information about the circumstances at hand, moral beliefs and views about critical and experiential interests. There is a tautology driving use of this shorthand: if patient decides on action A then the patient has decided on action A ($A \supset A$). Since I have been arguing that patient and the proxy are participating in the same cognitive process, it stands to reason each would come to the same conclusion. *A* cognitive process (in this case medical decision making) should end with *a* result. This is not exactly the case, however.

Even a single brain qua cognitive processer can come to different conclusions based on even slightly different data. The proxy and the patient are still separate persons and, as such, there will be some difference in how information is received and processed. What we are really looking for is a sameness of cognitive style. The proxy will reach a decision in *the same way* as the patient, even if the exact details of their decisions may differ.[129] Both an appropriate proxy

[129] I believe that the extended mind here eliminates any serious concern about the proxy making

and a virtual stranger could come to make identical decisions, but if the proxy's reasoning is closer to what the reasoning of the patient would be it is still the objectively better decision of the two. That said, I will continue to use the above shorthand, if only to simplify the process of making my arguments.

In short, the closer the thinking of the proxy is to the thinking of the patient is the degree to which we can say the proxy is transparent to the patient. Indeed, if the mind of the patient does extend into the proxy we are not dealing with different persons and examining the similarities of their thoughts, but instead have the thoughts of the patient revealed through an unconventional means. There will still be a difference in thinking as the two are separate entities but the extended cognition that that exists between the two will mean there is a great deal of isomorphic overlap. "Don't turn to scripture" for example, means the same thing no matter which of the brains in the group mind is considering the imperative.

A Constant in One's Life

There are a couple of different meanings of the word 'constant' which need to be looked at here. On the one hand we need to consider the longevity of the relationship. Generally speaking, long and close relationships lead to the better/best proxies. But we can't always have both and longevity and immediacy nor are these always indicators of the right kind of relationship. Immediacy of the relationship is important, but there may be reason to give added weight to shorter term relationships which the patient is actively involved in over longer relationships which have become (for lack of a better word) passive, yet not quite estranged. Likewise, immediacy of a relationship does not guarantee the patient and the proxy qualify as a

---

decisions radically different from the patient. Then proxy might let the patient linger in a PVS longer than she might wish or, in tossup, where costs and burdens are about equal, choose option A when the patient might have just as likely chosen option B. It is

unified cognitive system. Prisoners spending decades in the same cell or siblings living at home well into adulthood might be very familiar with one another yet still not be ideal choices for a proxy.[130] Beyond this, a patient is able to designate any proxy she sees fit, regardless of the length or immediacy of the relationship, so long as the appointment was an autonomous decision.

In what follows I will look at both these important characteristics (longevity and immediacy of the relationship). In doing so I hope to give guidance for determining when a long relationship is the right kind of long relationship; ditto for immediacy. There is also the matter of justifying the possibly short-term, immediate relationships and long-term yet passive.

Longevity

The chances for cognitive coupling are increased over time. Though this thought is not a ground-breaking discovery, it does nonetheless warrant some exploration. If a person speaks to someone every day for many years, there is a far better chance of isomorphic overlap between their brains than someone to whom she only speak to occasionally. What this over simplification forgets is that frequent interactions are not always equally meaningful. I might see the same barista every work day for 10 years but if our interactions are limited to small talk and drink orders the barista would serve as a good proxy only in so far as I need someone to order a drink for me or informing those around me of my feelings about the weather. Certainly, this is a *kind* of isomorphic overlap but not nearly enough to sustain an argument in favor of the barista as an appropriate proxy.

On the other hand, my best friend, Michael, is someone I speak to only a handful of times a year. We live in different cities and are busy with jobs and family lives. Michael and I have known each other for more than 25 years. In the past (high school) we saw and spoke to each

[130] This is not to say such circumstance can't lead to relationships well suited for proxy-ship.

other every day but have since developed a relationship which is not as immediate, yet no less intimate. Though our interactions are far less frequent than those with my hypothetical coffee guy, I feel supremely confident he and I have developed sufficient cognitive isomorphism to make him a good proxy (and vis versa) should the need arise. Our interactions have been at times trivial and juvenile at times,[131] but we have also shared a lifetime of experience and are as well acquainted with each other about as much as two people can be.

The longevity of a relationship is a barometer of the *probability* of isomorphic overlap. Just as smoking does not guarantee someone will get lung cancer, neither does time itself cause mental connectedness, only that the odds are increased. The right kind of long relationship is one in which a great amount of relevant information has passed from one brain into the other. The barista is in receipt of *some* information about me, but the information passed along is not very telling of me as a person. Although I cannot relay to Michael everything that goes on inside my head, no matter how much time we have, there is a great deal more bandwidth existing between us. Pieces of information overlap and interplay with other pieces of information. So much so he will, in some cases, be able to predict my response and behavior in a given, complex situation. At most, the barista might be able to guess I will want an iced coffee because it's hot outside.

Immediacy

People change. It is a poor, sheltered person who claims to be exactly the same person as he was as a teenager as he is twenty years on. Not only can our tastes in trivial matters such as music and food change, but more importantly our morals, politics, relationships and desires for

[131] Our wives would say it's a bit more than "at times".

our lives can change greatly over time, perhaps even a short period of time, under the right circumstances.

A person, in their youth and early adulthood may have been politically conservative and deeply religious, yet in later life liberalized his thinking and turned to atheism. These sorts of changes of attitudes and opinions will lead people from the person's past to come to different conclusions about the person's wishes than those who are close to him now. Since an advance directive is a document written in person's past (be it a few hours before surgery or many years prior) the possibility exists that the mind of the patient may have changed since it's drafting. It is the job of the proxy to tell us what the patient believes or desires now versus what was proclaimed in the past. The past directive and current beliefs of the patient may be in conflict. A potential proxy who is currently, actively involved in a patient's life is perhaps in a better position to know such a conflict exists and how to resolve it.

So, Do We Choose Longevity or Immediacy?

Neither the length or contemporaneousness of a relationship are on their own a sufficient criterion of choosing the right proxy. Given time and closeness, the minds of individuals are increasingly more likely to form the kinds of isomorphic overlap which makes someone a good proxy. But choices need to be made and the medical staff need to know who the right person is to make these choices.

The ideal proxy would be from a relationship which has a long history and is still ongoing, thus meeting both of these important criteria. It is here that we can most effectively discuss the notion of unusually interdependent couples. These are couples that, according to Clark and Chalmers "one partner's beliefs will play the same sort of role for the other as the notebook plays for Otto." (Clark and Chalmers, 1997). This definition alone is not sufficient to

set these couples apart from other coupled relationships (like the phone number example) but I have previously alluded to the further assertion:

> "What is central is a high degree of trust, reliance, and accessibility. In other social relationships these criteria may not be so clearly fulfilled, but they might nevertheless be fulfilled in specific domains. For example, the waiter at my favorite restaurant might act as a repository of my beliefs about my favorite meals (this might even be construed as a case of extended desire). In other cases, one's beliefs might be embodied in one's secretary, one's accountant, or one's collaborator." (Clark and Chalmers, 1997).

What makes a pair (or possible a group of three or more persons) "unusually interdependent" is the areas in which trust, reliance and accessibility are shared. I would like to be able to give a continuum of seriousness of decisions one faces in life, but such an attempt would be a fool's errand to be sure. That being said, there are at the extremes some decisions we can, with confidence, consider trivial and others of the utmost importance. Where to have lunch or what to watch on TV are pretty clearly trivial compared to whether or not to continue life-sustaining treatment or to "pull the plug" in hopeless cases.

There are degrees of connectedness between all persons with more than a passing relationship. But in order to rise to the level of being a truly 'interdependent couple' those connections will have to run deep and be of the utmost importance. One example of how we could begin to make the case that an unusual interdependence exists is whether or not a history exists of a couple making important decisions for one another (such as life or death or otherwise serious medical choices) and the other endorsing those choices ex post facto. If the one for whom the choice is being made endorses the decision (i.e.- specifically accepts the decision as the decision she would likely have made in the same circumstances, not merely finding it "acceptable") of the other time after time the so called 'proxy' is actually a revealer of the other's mind. This may strike the reader as little more than inductive reasoning based on past

evidence. However, if other criteria for the extended mind are met we might not be able to make this claim of simple induction so easily.

For example, when the option is available, I choose mint chocolate chip ice cream over other flavors, almost without exception. At a future date when mint chocolate chip is available my choice to have it is not traditionally what we would call an inductive decision (from my point of view). We can, of course, frame the choice in this way by describing the decision such as:

> In the past I have enjoyed mint chocolate chip ice cream every time I have tried it. I have had other flavors of ice cream at various times when mint chocolate chip was not available and by comparing these experiences I can conclude that I prefer the taste of mint chocolate chip as I recall having experienced it. Therefore, in the case before me I will induce that the selection of mint chocolate chip will maximize my enjoyment of ice cream in the near term.

It is entirely possible that this sort of process takes place, unconsciously, every time we make a choice. This, however, is not how we, ourselves, or others, commonly describe our decision-making progress in practice. If I choose mint chocolate ice cream at a particular instance and someone asks my wife why I did so she would say something to the effect "because he likes it best" and not "he has induced from past experience that this choice will create the best state of affairs in terms of flavor." My preference for mint chocolate chip ice cream is not an argument I make with myself, it is just a fact of preferences I have.

If the mind extends into the brains of others the same kind of reasoning holds and we do not need to resort to talk of inductive decision making. My preferences for what is a life worth living and when I would be better off dead are facts about me and my preferences, not continual inductive arguments. This desire can extend into the brain of another who is socially situated (relative to me) in the right way. My preference of mint chocolate chip ice cream has roots in the phenomenal experience of eating this flavor (as does my dislike of eggs and canned beets), but the fact "Michael likes mint chocolate ice cream more than any other flavor" is an isomorphic

fact which survives with complete fidelity across media. Whether I think this to myself, write it down or it is said aloud by someone who knows this to be true of me, the information is isomorphic despite its having phenomenal roots.

The same can be said for a desire to not live in a demented or otherwise severely compromised state. Whatever the reason may be for finding such a state undesirable, the fact that one finds this state undesirable is simply a fact about that person and this fact can be expressed in numerous ways. The partner in unusual interdependence is the most reliable means of revealing this fact.

An appropriate proxy need not rise to this level of interconnectedness, but it would be nonetheless the strongest kind of relationship possible in regards to deciding for others. These kinds of relationships would likely be rare. The degree of isomorphic overlap necessary to say that two otherwise separate persons participate in the same mental life and the one speaks for the other with complete authority in all things is not a relationship most people will experience. Familiarity and isomorphic overlap, like many human traits, exists on a spectrum. There are those people who know me extremely well and share a fair deal of overlap and others who I have never met for whom what overlap exists is either accidental, coincidental or trivial.

If the longevity and immediacy of relationships is considered when choosing a proxy, it is more probable we will find a proxy who is more a revealer of the mind of the patient. Insofar as a written directive is a copy of the isomorphic desires of the patient, so to would be the expressed decisions made by an appropriate proxy. As the situation stands now we treat a written directive as coming directly from the patient and decisions of the proxy as mediated in some way that makes those decisions open to greater suspicion. There will of course be times where the decisions of the proxy are suspicious but all in all there is no more reason for doubting

some proxies as compared to the written directive. Directives can be written under duress or in a state of undetected/undiagnosed incompetence or in a moment of weakness (which, if known, would invalidate the directive due to a lack of autonomy in its drafting). All decisions regarding medical care (in fact, any decision on any matter) are open to some form of doubt or other as to whether or not the decision really is what the decider desires, all things being equal.

Has Been Endorsed in the Past and there is Some Consequence to its Future Endorsement

A patient may select a proxy. So long as this appointment is an autonomous, competent decision this is as clear a case of endorsement we are to ever find. The selection of a proxy is, by definition, an endorsement which carries consequence; in some cases, grave consequences. The appointment of a proxy by the patient means this criterion for extended mind has been met.

If every competent person appointed a proxy and there was no additional controversy concerning the appropriateness of said proxy the matter here would be closed, but we do not live in such a world. In cases where no proxy was previously appointed by the patient we can look at past endorsements and attempt to apply them to future choices. Someone who has made important decisions on the patient's behalf in the past and had those choices retroactively endorsed is a better candidate, all things being equal, then someone who has not made such choices or received such confirmation.[132] These past decisions need not be matters of life and death. From time to time, in any close relationship, one party or the other will have to make an "executive decision" on behalf of the whole. These decisions might be the same as what the partner would have chosen or not. Either the non-deciding party endorses the choice (explicitly or tacitly) or "lodges a complaint" with the executive. In either case the decision maker has

[132] Again, we face the possible criticism of relying on inductive reasoning and substituting that for the actual expressed desires of the patient. This concern was addressed in the previous section and as such will not be repeated here.

gained insight into how the non-deciding party would have chosen. If I tend to agree with choices made by my brother more often I do those made by my sister, my brother is the more appropriate proxy for decisions of those types. The kinds of executive decisions most likely to be made throughout the life of a relationship are often practical matters; should I go with contractor A or contractor B? Would my child like a Tonka truck or a Barbie for her birthday? At which restaurant should I put in a reservation? There may be reason for disagreement between the parties involved but the stakes are not always going to be terribly high.

Medical decisions are of a different kind. There are practical concerns, of course, but more importantly these are sometimes choices about what one values and what we value is not going to be considered by everyone to be equally practical. If I value my biological life, regardless of my mental state or prognosis for recovery I will see being kept alive on life support to be imminently practical, whereas the hospital will see this as a poor use of resources. A proxy need not share her charge's values but she must be aware of and willing to respect them if she is to be a good proxy. It almost goes without saying that the closer the proxy's choice is to matching my values the greater likelihood we would give our theoretical endorsement to that decision.

Non-person Proxies

However, a proxy does not even need to be a person. Consider, the Jehovah's Witness patient in need of a blood transfusion discussed earlier (an example borrowed and adapted from Nelson, 2003). If in the past she has deferred to the tenets of the church in making decisions, even if she had only learned of them (by this I mean if she, consciously and intentionally, changes her mind immediately after learning of a change in church doctrine or upon learning

new facts) we are on firm ground if we defer to what the church says if she is not able to self-report on her wishes.

Religious doctrines are, perhaps, the most likely or common examples of a non-person proxy but they are not the only such possible proxy. Say for example I live my entire life by making all decisions by flipping a coin. If I do this long enough, and for serious choices (should I get married, should I get a flu shot, should I pull the rip cord on this parachute) it is acceptable, perhaps even incumbent, upon medical personnel to make life or death choices on my behalf using this method. It's an odd, risky, even dangerous way to make decisions of life and death, but it is consistent with my values to defer to a coin in this case. How a person makes choices throughout her life gives strong inductive evidence that she would do the same at future points, were she able.

Is this the Extended Mind or Just Good Qualities for a Proxy?

Is all this just a method for selecting a proxy or does the extended mind really play a role in this? It seems we could apply the content of Clark and Chalmers, and Levy's criteria and never bring the extended mind into question. But by doing so I believe make a weaker case for the selection of a proxy. Examining the criteria alone, apart from the extended mind theory from which they arose, makes the criteria arbitrary and in need of argument or justification. We could have just as easily selected a different set of criteria for selecting a proxy. Such as:

- The proxy is a blood relative of the patient
- The decision of the proxy is in line with Christian scripture
- The proxy does not stand to benefit financially from the death of the patient.

Not all collections of criteria are going to be equally valid or useful. An argument has to be made to justify one's choice of criteria. The argument in favor of the extended mind criteria

make a strong case in favor of their selection. If the extended mind theory is true, the criteria to test for it have a great deal of force not available to other sets of criteria. The arguments in favor of the extended mind thesis make these criteria hang together coherently. This dissertation has been an attempt to show the validity and applicability of these criteria to medical cases.

With few exceptions (see Varelius, 2006 for one such example) we respect the wishes of the autonomous, competent patient even if we believe these choices are not, from our point of view, in that patient's best interest (and also because we respect the value of a person living her own life). The reason is that, as the orthodoxy stands now, a person is that best judge of what is best for her. This is in large part due to the fact that (again as the orthodoxy stands) persons have privileged access to the contents of their own minds. Only the person to whom the mind belongs can know what is "going on" in that mind.

The extended mind changes this. The mind is not wholly closed off from the rest of the world. If the mind leeches out into the world and into other people, then outsiders cannot only see it but perhaps take part in that mental life. Isomorphic mental content is not located solely in the brain of the content's originator. Just as my wife can tell you, authoritatively, that I like mint chocolate chip ice cream so too she can make my medical wishes known should I be unable. Those parts of mental content on which we base our actions (i.e. – the words that make up the instructions) are the same across media (in the brain or on paper) and the phenomenal qualities which remain internal to the mind no longer have causal power. Their causal power halts once the desire is drafted in the brain in an isomorphic form. This phenomenal content has the capacity to exert causal power again, which manifests in the fact that, over time, we sometimes change our minds about our desires, wishes, likes and dislikes. Wayne, as a child had the phenomenal experience of liking cheese which could be expressed isomorphically as the

statement “I like cheese” or “Wayne likes cheese”. Later in life as Wayne’s phenomenal experience of cheese changed so too did the form of isomorphic expressions (“I don’t like cheese” or “Wayne does not like cheese”).

The most appropriate proxy is one who reveals the patient’s decision in an isomorphic form just as the patient herself would. This being the case, the proxy is, in a real sense, sharing a mental life with the patient because of the isomorphic overlap between the two otherwise disparate minds. This is a much stronger claim which can only be made if the extended mind thesis is true or assumed to be true.

Applying the Extended Mind Criteria to Advance Directives

Just as we applied the criteria of the extended mind to proxy decision makers, so too can we apply them to advance directives. In many ways the reasoning runs parallel in both cases, but there are some slight differences in application of the criteria since we are dealing with a (mentally) inert document drawn from the phenomenal and isomorphic content of the drafter’s mind rather than the (mentally) active mind of a potential proxy.

Readily Available and Easily Accessible (Transparent in Use)

In the context of advance directives there is significant overlap between the applications of these criteria, so it is expedient to handle them together. What is true of one will often be true of another.

We are not so much concerned with how ready at hand the actual document is. I doubt there are many (though surely there are some) people who carry their directives on their person at all times as some might wear a medic alert bracelet. But whether these documents are on file in a hospital, lawyer’s office of home filing system or on a computer, let’s assume the people

who need to get to the document(s) can get to them. Given that, what do accessibility and transparency mean?

These refer to the clarity of the document and its applicability to the situation the patient finds herself in. It can be assumed that the patient has it in her best interests to make the meaning of the document as clear as possible. Shades of meaning might exist and indeed might be unavoidable given the medium of language. A clear statement of intent within the appropriate context is going to be a much better approach than reading between the lines.

The more tortured the interpretation the less likely it is part of the wishes of the patient. Tortured interpretations may, in the end, be reflective of the patient's wishes but, as with many things simplicity is to be preferred. The patient has written a document of grave importance. It might well be a sign of incompetency if the only way the document makes sense is through protracted, "creative" interpretation.

A Constant in One's Life

Though this is an important criterion for the extended mind (it takes time to fully incorporate an external tool into one's mind proper) an advance directive does not require a long engagement period before the internal/external coupling is completed. Not all advance directives are intended to be a permanent part of a patient's medical records. The document could be drafted in anticipation of a scheduled surgery, in which case after surgery and recovery the directive might not apply. In addition, competent patients can rescind advance directives at any time. This may weaken the importance of this criterion but what makes this weakening acceptable is the strength the relative endorsement criterion discussed in the immediately following section of this chapter.

A more appropriate understanding of the constancy criterion would be to frame it as 'a constant for the relevant period of time.' If the advance directive is one meant to be of limited lifespan (say to cover a particular surgery or medical procedure) we can say the directive is a constant in that person's life since the span of time between drafting and the need to implement may be very short (perhaps as short as a couple of hours). While unconscious (since the directive would not be in force until the patient was unconscious or otherwise unable to make medical decisions) a patient is not able to change her mind and the directive is all the record of her mental life which remains accessible to medical personnel and loved ones.[133]

Cases of longer lasting directives (meaning those drafted with no particular event in mind but drafted as a safeguard to protect the patient's wishes in the future) are a trickier matter. Though it is not a new problem today we are seeing more clearly than ever the results of a society which preserves documentation well past its applicable life span. As the saying goes, once it is on the internet it is on the internet forever. I doubt many people hold the exact same beliefs on complex issues now as they did decades previous. Perhaps the big picture remains the same but certainly reflection will create nuance to a person's thinking.

If a patient was an atheist thirty years ago but today has found religion, we might do her a great disservice by implementing an advance directive drafted under her previous belief system. This remains a problem only as long as an appropriate proxy is not found. If a decades old directive is on record should it be abided by if no appropriate proxy is present to tell us otherwise?

[133] However, the continued existence of the directive after it has expired serves as a record of what the patient once thought though she my no longer think the same thing.

If there is reason to suspect that the drafted directive as written is no longer applicable to the patient and no obvious, appropriate proxy is to be found, we are barely farther ahead than not having any directive at all. This being the case, we will need to look at the directive within the context of the person's life and life history. It is possible, I suspect, that a close examination of a person's environment might yield enough information about that person to allow even strangers to make an informed choice for the patient with a very good chance of acting in accordance with the patient's wishes. See this dissertation's appendix for more on this possibility.

Has Been Endorsed in the Past and there is Some Consequence to its Future Endorsement

As was the case with selecting a proxy the drafting of an advance directive, by its nature, is an endorsement of future consequences. What is different is that the written directive is static in nature. The words on the page are always those words on the page. Open to interpretation, yes, but nonetheless unchanging qua word on paper. We cannot go back to the written directive for clarity with any hopes higher than to refresh our memories about what is said. There won't be something written down tomorrow that was not there yesterday. With a proxy, especially a proxy participating in the same mental life as the patient, we do have this luxury. As the circumstances of the patient and the world change, a proxy can formulate and inform us of changes in how the patient would respond in light of new facts. Not true of the directive.

As time passes the chance for a change of heart given changing circumstances increases. In the extreme a directive might address a condition which was incurable/untreatable at the time of its drafting, but which has since become a routine, relatively harmless procedure at present. It's not a stretch to believe the patient would accept such a new treatment, but the directive may give us no indication one way or the other of what the patient would in fact decide. This being so, I argue that the directive, as written, stands regardless of changes in circumstance unable to be

addressed by an appropriate proxy. As was the case with Odysseus and his desire to not succumb to the siren's song, we here have to default to the last known and maybe only remaining, competent mental state applicable to the patient. It is a necessary component of the competency to draft an advance directive that there is the understanding that the document will apply to one's self at a future time (since that is the overall purpose of the document in the first place). A properly drafted directive should take into account the possibility of changes in future circumstance and the medical state of the art. An expression of *why* the drafter has made the choices she has could go a long way to aiding those charged with making decisions for the patient.

Extended Mind or just Qualities of a Good an Advance Directive?

The answer is yes on both counts. A good advance directive is, in fact, part of one's extended mind. It is not an either/or proposition. As has been previously argued, an advance directive is an attempt by the drafter to project her wishes into the future through an observable, mental state. The directive is a persistent act of will, not mere guidelines or suggestion left for the future to sort out. The directive is an isomorphic duplicate of the patient's wishes and as such it is just as good as asking the patient what she desires.

The same reasoning holds here as it does in the above defense of proxy decision makers using the extended mind. The criteria for being part of an extended mind make an excellent set of criteria for advance directives and the defense of the extended mind thesis makes the criteria non-arbitrary and systematic. If an advance directive is part of a person's mind, that is to say is a mental state of that person, we are a long way toward resolving the K-PADS outlined in chapter 1.

Conclusion

The unknowns in our knowledge of the mind leave the door open to unorthodox thinking on the subject. The extended mind thesis is one such unorthodox view of the mind, but this makes it no less plausible than any number of theories of the mental. And, in fact, reflecting on and refining the extended mind even only *as if* it were true opens additional doors in seemingly unrelated fields, such as ethics, and allows us to seek answers to difficult problems in new ways. It was my hope when I began this project to add new avenues of discussion to both the debates about extended cognition and decision making for others. More specifically, I wanted to explore the connections between these two debates in the hope of answering some of the more persistent questions and doubts found in each of these areas. I will leave it to the reader to ultimately decide if I was successful.

No one serious about the human mind would doubt that it exploits its ability to interact with its environment to lessen cognitive burden or speed up cognitive processes. Very few can easily divide 5692 by 12 easily without the aid of longhand long division. We're even faster when we use a pocket calculator. This is why Andy Clark refers to members of our species as "Natural Born Cyborgs" (Clark 2003 and elsewhere). Our very nature, composition, essence or what have you, is to be a being which uses the environment to aid those most human characteristics: higher thought, rationality and cognition.

Presumably humans have been extended in our minds as long as we have been humans. It is probably no accident that nearly all societies, when they developed a counting system did so in base 10 (to corresponds with a standard compliment of fingers). As our need for faster and more complex computation and cognition grew, so did the ways in which we learned to exploit our environments, and each other to get the job done. Language made it possible to pass along more

complex ideas. The written word let us pass these ideas through time. The printed word allowed that information to spread farther and faster. As we freed our minds from certain tasks we freed them up to engage in others. Over time we needed more information to get through our day-to-day lives than we could hold in our heads. Address books and calendars and later "all-in-one devices" such as electronic organizers (and eventually cell phones) helped keep names, dates, phone numbers and times readily at hand without creating undue burden.

As humans passed from a life that was "solitary, poor, nasty, brutish and short" to the Renaissance, the Enlightenment, the Scientific Revolution and the Computer Age our cognitive needs changed as did our relationship to other people. We are involved in an ongoing process in which we are becoming more aware of, empathetic to, and morally concerned with others. Our moral codes have gone from "do unto others" to treaties sometimes hundreds of pages long. Our legal codes have grown from "10 commandments" to litigious systems spanning nations and centuries. It is not difficult to surmise our "natural born" ability to co-opt our environment to aid our cognition should find a foothold in our moral lives.

Though the *idea* of sustaining a human life through artificial/mechanical means dates back at least as far as 1530 when Paracelsus used a simple fireplace bellows to inflate to lungs of recently deceased persons, it wasn't until the early 20th Century that actually keeping a person alive with the aid of machine was a real possibility (Snider, 1989). Though deciding for others became an acute problem at this time, no longer did the stoppage of the heart, lungs or brain mean the stoppage of all other organs would shortly follow. The death of the body and the death of a person were temporally closely tied together. For nearly a century we have had the ability to replace some critical functions of the body and thus extend life past the point of a patient's personhood.

Proxy decision makers have always been needed when a patient could not decide for herself. But as life-sustaining treatment improved, so did this need grow. Later we developed a legal instrument, the advance directive, to allow a person to make her own wishes known were she to become unable to express them. Though proxies and directives are implemented to solve practical and legal problems concerning end-of-life decisions, I have argued here they are more than that. Our relationships with others and our written directives are an expansion of our capacity for mental extension. To make one's wishes known to others is to further expand our mental life and autonomy beyond our own corporeal lives. It is not a stab at immortality as an artist or philosopher might seek by leaving her work behind, but it is our using our status and natural born cyborgs to keep our lives under our own control for as long as we deem necessary or is possible. If autonomy is a principle of the highest importance, it stands to reason we should respect the autonomy of patients no matter the form that autonomy makes its desires known to us.

APPENDIX

Problems of Absence

In my introduction I divided the problems of advance directives into three categories: problems of personal identity, problems of interpretation and problems of absence. The first two categories were discussed in chapters 3 and 4 respectively. It is in this appendix that I will begin to address the third.

Why not just add an additional chapter? In fact, this *was* the plan. It was not long, however, before I realized this topic could get out of control very quickly. Instead of looking at a discreet directive or evaluating the appropriateness of one possible proxy over problems of absence will require an examination of a person's life and environment as a whole; a considerably larger domain of discourse. This topic alone could be the subject of its own book at least as long as this dissertation. Rather than add another 200 pages and several months to this process, I have instead opted to summarize how the extended mind *might* be able help to resolve these problems of absence and leave the details for a future project.

What follows is part literature review and part speculation. There is some promising research concerning how Alzheimer's patients function better in familiar versus unfamiliar environments. The results of these early investigations indicate there might be something beyond practice and routine driving this increased performance. From here I will speculate on how we might be able to determine a person's thoughts and desires through an examination of her environment. In effect, if we are "natural born cyborgs" as Andy Clark suggests (Clark, 2003), we might leave traces of our minds scattered throughout our environment just by living our lives. These mental leavings create a "paper trail of the mind" which, like other instantiations of the extended mind, are observable from outside a first-person point of view.

Types of Absence

The first step here is to reiterate and separate the two different reasons for absence of an advance directive. In doing so, a large portion of directive-less patients will fall outside the scope of this dissertation. The first are cases where no directive is present because the patient was never competent enough to draft it (problem number 6 by the ordering in chapter 1). These are cases of the young, the cognitively disabled[134] and the otherwise incompetent. The second cases are those in which a patient was competent but nonetheless never drafted a directive for one or more of any number of reasons (problem number 5). Of the first group I will have very little to say. A young child (say an infant) has not developed the mental faculties to have desires about how she wishes her life should go. The same can be said of the severely cognitively disabled. The extended mind thesis is of no help in these cases and as such I will have nothing more to say about them here.

Of the second group I only have a few remarks I wish to add to the conversation given the psychological research that exists on the relationship between the person and her environment. In the cases of problems of interpretation and personhood, the extended mind thesis gives us reason to believe that the mind of the patient is observable to outsiders in a direct way which leads to more certainty about what the patient wishes. This is not the case with problems of absence. When a directive is absent there is no intentional, explicit mental effort left by the patient concerning how she wishes her treatment (or non-treatment) to proceed. No written instructions and perhaps not even an acceptable proxy. In such cases the extended mind thesis gives us additional guidance as to how the patient might choose by looking at the interactions

[134] Cognitive disability per se does not preclude one from being competent to draft an advance directive. Such disability to a severe enough degree would.

between the person, her environment and her actions. We are still making a best guess based on the available evidence, but the extended mind can give us more confidence in how we interpret the evidence.

The Competent but Directive-less.

We face similar problems to those found in the previous section concerning interpretation of advance directives with the added hurdle of a lack of external (from the point of view of the patient) guidance. If a proxy exists we need only defer to her in order for a decision to be made. Ideally the proxy will fit the criteria for extend cognition. This being the case, the directive-less yet competent patient presents no more challenges than have already been addressed previously in this dissertation.

The challenge to be tackled here is what to do when there is no proxy which shares in the mental life of the patient in the strong sense discussed in an earlier section. Are we limited to the best guesses of a less than optimal proxy, or worse yet no appropriate proxy at all? I will argue here the answer is no, but at the same time admit that much of what follows is very much open to doubt, speculation and misunderstanding. What I purpose is not perfect, but it is, perhaps, better than nothing.

Cues for what I propose come from various fields, but chief among them is criminal psychology, that is to say criminal profiling. Early on I distanced criminal profiling from "racial" profiling; the difference being that criminal profiling is the loose science that given the evidence at hand what can we learn about the perpetrator of a crime (given the way the cut traverses the neck the assailant was likely left handed, etc.) versus racial profiling which is the belief that given evidence in the past what might we predict about a person before any crime has taken place (since a large number of mass shootings in the United States have been perpetrated

by non-mentally ill, white, males (Follman, 2015), we should subject all such persons to added scrutiny.

What can we learn about a person from their environment

Examine what sort of food is in your kitchen. What can this teach us about your eating habits? Do you buy organic produce? Do you by brand name or off brand cereal? Is your fridge full of ready-to-eat food or would some serious preparation be necessary to get a meal to the table?

This is, admittedly, an inductive process. If you inspect my fridge on some days you will see some half empty bottles of salad dressing and not much else. You may be lead to the conclusion that I either eat all of my meals out or that I am in desperate need of a trip to the grocery store. But as with all good inductive investigations, there is more to go on than contents of the icebox. I have a large collection of cookbooks, high-quality (and well used) cookware, a well-stocked pantry and a spice rack which rivals a Turkish souk. Collectively the evidence favors someone who cooks. This is inductive, but the odds of learning something true about me by examining my entire kitchen are much better than a blind guess.

We leave pieces of our minds all over the place. More so today in a world where we are constantly adding to our data footprint. Our phones tell us who we talk to, GPS tells us where we have been, and our apps tell us how we try to be productive and how we spend our idle time. We have already seen an example of the extended mind which fit the bill for determining a patient's wishes in the absence of an advance directive: our Jehovah's Witness patient from chapter 4. It was argued that this patient's past acceptance (nearly unreflective acceptance) of the tenets and beliefs of the church leads to the possibility of her changing her mind without knowing it. That is to say if the church has a change of heart and allows blood transfusion we

are choosing as the patient would when we give her a blood transfusion which she is unable to consent to.[135]

[Alzheimer's sufferers in St. Louis] were a puzzle because although they still lived alone, successfully, in a city, they really *should not have been able to do so*. On standard psychological tests they performed rather dismally. They should not have been unable to cope with the demands of daily life. What was going on?

A sequence of visits to their home environments provided the answer. These home environments, it transpired, were wonderfully calibrated to support and scaffold these biological brains. The homes were stuffed full of cognitive props, tools and aids. Examples included message centers where they stored notes about what to do and when; photos of family and friends complete with indications of names and relationships; labels and pictures on doors; "memory books" to record new events, meetings, and plans; and "open storage" strategies in which crucial items (pots, pans, checkbooks) are always kept in plain view, not locked away in drawers. (Clark, 2003)[136]

Clark believes one plausible explanation for this phenomenon is the extended mind. If a person has spent time (perhaps a lifetime) in an environment it is possible the things in that environment have been incorporated into her cognitive system just as Otto incorporated his

[135] We must be careful not to paint with too broad a brush, else we run the risk of stereotyping or turning a person's most deeply held religious beliefs into a caricature of who they are. Despite the proclamations of the Catholic Church that no woman should use any birth control, between 82% and 98% (the numbers represent the high and the low results of various polls conducted between 2001 and 2014) of American Catholic women admit to having used some kind of birth control in their lifetime. Not all Jews keep kosher and not all Muslims pray toward Mecca five times a day.

[136] Clark twice (Clark, 1997 and Clark, 2003) cites [D Edwards, C Baum, N Morrow-Howell Home environments of inner city elderly with dementia: Do they facilitate or inhibit function? Gerontologist 34 (1994)]. After repeated searches and correspondence with Carolyn Baum (the second author) it has been determined that no such article exists (per se). What Clark appears to be referring to would properly be cited as: ["Abstracts." The Gerontologist 34. Special Issue I (1994): 1-402.]. This is a program of abstracts and speakers from the Gerontological Society of America 47th Annual Scientific Meeting November 18-22, 1994 Atlanta, Georgia.

I gather this is a talk given by the cited authors. I cannot find any one paper which discusses the topic so nicely summed up by Clark (2003 and quoted above). These themes are found in the following sources: (Baum, et al., 1993), (Baum and Edwards, 1993), and (Edwards et al, 1999).

notebook. The parallel being: if it is in Otto's notebook, Otto must have thought it was important, if something in a person's environment was made so by the person, it must have been important for it to be that way.

The average American weighs 175 pounds, and the human brain weighs in at an average of 3 pounds.[137] Though humans have the largest brain-to-body mass ratio of any animal[138] (Sagan, 1986), the brain is still a mere 1.71% of the human body by mass. Despite its relatively low mass, the brain is far and away the most energy intensive organ in the body. Of all the energy we take in as nourishment and oxygen the brain uses 20% of it (Swaminathan, 2008). The reason is that cognizing is hard, ceaseless work.

We are familiar with the difficulties and limitations of cognition. We write notes to help us remember important information. We use calculators to add sums and do division. We wear watches to track the time. We ask the person next to us to "remember this number".[139] All of these processes have at least two things in common: first, they have been argued to be part of our minds (should certain circumstances be met) and second, we intentionally adopt them into our cognitive apparatus. Given, however, that most of what our brain does happens at a subconscious (or non-conscious) level, it stands to reason that our sub/non-conscious mind would seek out ways of lessening its cognitive burden.

A thought experiment purposed by John Case (first mentioned in Chapter 2) can help illustrate this point. In the thought experiment we are asked to consider a patient (call him John) with ever worsening dementia. In order to compensate for his failing mental functions, John creates a

[137] That's 79.37 and 1.36 kg, respectively, for the metric minded.
[138] Save the pygmy shrew which has almost the exact same ratio. However, it's brain is overall so small it almost doesn't have enough processing power for basic housekeeping.
[139] I promise, this is the last time (in this dissertation) I will use this example.

robot to assist him. As John loses function, the robot takes over that function until nearly all activities are taking place in the robot rather than John's brain. Case means this to illustrate how the mind can extend over substrates outside the skull by showing the extension process taking place over time with a person's active involvement.

Now, replace the robot with John's environment. As John believes he is becoming more incapable of running his life "from the inside" he starts to change his environment to ease the increased cognitive burden. He leaves notes lying about, keeps important documents close at hand and anything else he thinks might be of assistance. Just as the mental functions taking place in the robot originated, at some point, in John's brain, so too do the environmental changes now working as part of John's extended mind.

Now, make one further change, John doesn't have dementia, he's just some guy living his life interacting with his environment, making it more amenable to his existence as a being with a limited cognitive capacity (as we all are). John will both consciously and unconsciously manipulate and change his environment, to suit his mental needs. Books that are on a shelf in a well trafficked room hold more significance to the owner than those stored in boxes in the attic. Further, worn books with marginalia are better known by the reader. Photos visible from a person's favorite chair (which we might be able to determine by examining the upholstery) are of people and places dearer to the one who favors sitting there. If we could reverse engineer how a person's environment came to be in the state it is in, we would be closer to understanding that person, perhaps in morally significant ways.

We engage in something like this reverse engineering every time we induce something about a person based on evidence at hand. For example:

- I think this guy likes unicorns, because his living room is filled with unicorn themed stuff.

- He has got to be an MSU fan. His den is decorated in green and white and there are lots of pictures of Sparty, The Breslin Center and Spartan Stadium.

From these observations we can draw some conclusions about the likely future behavior of these persons. In the first case, if the unicorn lover were given a choice between a new blacklight unicorn poster and a poster depicting the "Creation of Adam" by Michelangelo we could be confident that the former would be the choice. We could also safely assume the alleged MSU fan would choose to watch the Michigan State Spartans play the Michigan Wolverines game instead of watching the Scottsdale Community College Fighting Artichokes face off against the Campbell Fighting Camels.[140] But this simple inductive inference is a long way off from determining, with a high degree of certainty, someone's end of life wishes. Yes, we are learning something unspoken about a person's mental life but what we're learned is trivial.[141]

More seriously, this induction of a person's motivations for future behavior has been used in at least one area of serious study and inquiry: criminal profiling. By examining the evidence at a crime scene and comparing this to past findings in similar cases, investigators are able to determine, with accuracy better then chance, what is likely to be true of a perpetrator. Such as:

- Given the depth and direction of the cut we believe the killer to have been taller than the victim and left handed.
- There are many more stab wounds (or bullet wounds) then would be necessary to kill the victim. This indicates the perpetrator's rage with the victim, or someone significantly similar to the victim.
-

So not everything we can learn about a person's interaction with her environment is trivial. Knowing that a person felt rage toward a victim tells us something significant about his motivations for taking serious actions. How we can apply this to the directive-less patient is still

[140] Yes, these are both real teams, but a match-up is unlikely. The 'Chokes represent a community college and the Camels are a NCAA division 1 team.
[141] That said, I know (and am related to) some serious Spartan fans. I could foresee MSU figuring prominently in some of their most important life decisions.

a bit of a puzzle, but perhaps the idea is promising given some previously mentioned empirical facts and simple suppositions based on those facts, about people's interactions with their environments.

- Humans, consciously and subconsciously, use their environments to lessen cognitive burden when possible. (Clark, 2003)
- Our brains do all kinds of things without our being aware of them. (Parfit, 1984 and Dennett, 1991)
- Given the sheer volume of external information available to the brain relative to its energy consumption and processing power, it is not implausible to assume the brain is changing its environment in way it finds useful and doing so without our conscious awareness. (Swaminathan, 2008 and Clark and Chalmers, 1998)
- Alzheimer's patients perform better than expected at tasks when in familiar environments (like their own homes) than standard competency tests indicate they should. (Baum, 1993 and Clark, 2003)

In order for this reverse engineering to become a useful tool in the medical ethics debate at hand, a great deal more study is needed in areas such as psychology, neuroscience, sociology, behavioral science and perhaps even interior design. Only the first baby steps in this direction have been taken but I believe it will prove to be a fruitful endeavor should others in these disparate areas choose to take further steps along with me.

BIBLIOGRAPHY

BIBLIOGRAPHY

Adams, Douglas. *The Hitchhiker's Guide to the Galaxy*. New York: Harmony, 1980. Print.

Adams, Frederick and Aizawa, Kenneth. The Bounds of Cognition. Wiley-Blackwell. 2008

Adams, Frederick and Aizawa, Kenneth. "Defending Non-Derived Content". Philosophical Psychology. Vol. 18 no. 6 December 2005

Adams, Frederick and Aizawa, Kenneth "The Bounds of Cognition". Philosophical Psychology vol. 14 no. 1, 2001

"ALS Association." *ALSA.org*. Web. 10 Oct. 2016.

Anthony, Sebastian. "Can a Loud Enough Sound Kill You? | ExtremeTech." *ExtremeTech*. Feb. 4, 2014. Web. 04 July 2016.

Applebaum, Paul S. and Grisso, Thomas. Assessing Patients' Competence to Consent and Treatment. New England Journal of Medicine. Vol. 320 no. 11 pgs. Mar. 1989

Aristotle, and W. S. Hett. *On the Soul ; Parva Naturalia ; On Breath*. Cambridge, MA: Harvard UP, 1957. Print.

Arras, John D. "The Severely Demented, Minimally Functional Patient: An Ethical Analysis." *Journal of the American Geriatrics Society* 36.10 (1988): 938-44.

Austin, John L., J.O Urmson, and Marina Sbisà. How to Do Things with Words. Cambridge (MA): Harvard U, 1975. Print.

Avramides, Anita. Other Minds. Routledge. 2001

Baum, C., D. F. Edwards, and N. Morrow-Howell. "Identification and Measurement of Productive Behaviors in Senile Dementia of the Alzheimer Type." *The Gerontologist* 33.3 (1993): 403-08.

Baum, C., and D. F. Edwards. "Cognitive Performance in Senile Dementia of the Alzheimer's Type: The Kitchen Task Assessment." *American Journal of Occupational Therapy* 47.5 (1993): 431-36.

Beauchamp, Tom L., and James F. Childress. *Principles of Biomedical Ethics*. New York, NY: Oxford UP, 1979. Print.

Block, Ned. "Troubles with Functionalism" Minnesota Studies in Philosophy of Science. Vol. 9. Pgs 261-325. 1978

Block, Ned. “Psychologism and Behaviorism”. The Philosophical Review vol. 90 no. 1 pgs. 5–43. 1981

Bonner, Stephen. “Are advance directives legally binding or simply the starting point for discussion on patient’s best interests?” Analysis. Vol. 339, pp.1230-1234. Nov. 2009.

Boyd, Robynne. "Do People Only Use 10 Percent of Their Brains?" Scientific American. N.p., 07 Feb. 2008. Web. 30 May 2017.

Bradley, F.H. “*In What Sense are Psychical States Extended”* Mind. New Series, vol. 14 April 1895. pgs 225-35

Brock, Dan W. "What Is the Moral Authority of Family Members to Act as Surrogates for Incompetent Patients?" *The Milbank Quarterly* 74.4 (1996): 599.

Brock, Isaac, et al. “Blame It on the Tetons.” *Good News for People Who Love Bad News*.

Buchanan, Allen. “Advance Directives and the Personal Identity Problem.” Philosophy and Public Affairs. Vo. 17 n. 4. Pp. 277-302. 1998.

Buchanan, Allen E. and Brock, Dan W. *Deciding for Others*. N.p.: Cambridge UP, 1990. Print.

Buford, Christopher. “Advancing and Advance Directive Debate.” Bioethics. Vol. 22 n. pp.423-430. 2008

Callahan, Joan, C. “On Harming the Dead”. Ethics. Vol. 97, pp. 341-35, January 1987.

Cartolovni, Anto, and Antoniog, Spagnolo. "Ethical Considerations regarding Head Transplantation." Surgical Neurology International 6.1 (2015): 103. Web.

Case, John. "Off-Loading Memory to the Environment: A Qualitative Example." Mind and Machines 14 (2004): 387-89.

Chalmers, David John. The Conscious Mind: In Search of a Fundamental Theory. New York: Oxford UP, 1996. Print.

Churchland, Paul and Churchland, Patricia. “Could a Machine Think?” Scientific American vol. 262 n. 1. Pgs. 32-39. January 1990

Clark, Andy. Natural Born Cyborgs. Oxford University Press. 2003

Clark, Andy. Supersizing the Mind. Oxford University Press. 2008

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

Clark, Andy and Chalmers, David. "The Extended Mind." Analysis vol. 58, no. 1, pp. 7-19, January 1998

Cole, David. "Statutory Definition of Death and the Management of Terminally Ill Patients Who May Become Organ Donors after Death." Journal of Medical Ethics. Vol. 18 pp. 26-30.

Connors, Alfred F. Jr, MD; Neal V. Dawson, MD; Norman A. Desbiens, MD; et al. *A Controlled Trial to Improve Care for Seriously III Hospitalized Patients: The Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments (SUPPORT)*. JAMA. Vol. 274. N. 20. November 22/29, 1995

Curd, Patricia, and Richard D. McKirahan. *A Presocratics Reader: Selected Fragments and Testimonia*. Indianapolis: Hackett Pub., 2011. Print.

Dawkins, Richard. *The Selfish Gene*. New York: Oxford UP, 1978.

Dennett, D. C. *Brainstorms: philosophical essays on mind and psychology*. MIT Press, 1979.

Dennett, Daniel. "The Myth of Original Intentionality. In K. A. Mohyeldin Said, W. H. Newton-Smith, R. Viale & K. V. Wilkes (Eds.), Modelling the Mind. (pp. 43–62). Oxford University Press (1990).

Dennett, D. C. *Consciousness Explained*. Boston: Little, Brown, 1991. Print.

Dennett, D. C. *Kinds of Minds: Toward an Understanding of Consciousness*. New York, NY: Basic, 1996. Print.

Descartes, René, and Donald A. Cress. *Meditations on First Philosophy: In Which the Existence of God and the Distinction of the Soul from the Body Are Demonstrated*. Indianapolis: Hackett Pub., 1993. Print.

Devlin, A. M. "Clinical outcomes of hemispherectomy for epilepsy in childhood and adolescence." Brain, vol. 126, no. 3, Jan. 2003, pp. 556–566

Dresser, Rebecca. "Life, Death, and Incompetent Patients: Conceptual Infirmities and Hidden Values in the Law," Arizona Law Review, 28:373-405, 1986.

Dresser, Rebecca. "Missing Persons: Legal Perceptions of Incompetent Patients." Rutgers Law Review. V. 46. N. 2. Pg. 609-719. 1994.

Dresser, Rebecca. Dworkin on Dementia: Elegant Theory, Questionable Policy. Hastings Center Report. V. 25. N. 6. Pg 32-38. 1995

Dworkin, Ronald. *Life's Dominion: An Argument about Abortion, Euthanasia, and Individual Freedom*. New York: Vintage, 1994. Print.

"The Einstein Approximation." *The Big Bang Theory*. CBS. 1 Feb. 2010. Television.

Edwards, Dorothy F., Carolyn M. Baum, Marie Meisel, Marilyn Depke, James Williams, Teresa Braford, Nancy Morrow-Howell, and John C. Morris. "Home-based multidisciplinary diagnosis and treatment of inner-city elders with dementia." The Gerontologist 39, no. 4 (1999): 483-488.

Evans, Allan Stewart, Riley E. Moynes, and Larry Martinello. *What Man Believes: A Study of the World's Great Faiths*. McGraw-Hill Ryerson. 1973 Print.

Feinberg, Joel. Harm to Others: The Moral Limits of the Criminal Law. Oxford University Press. 1984

FitzGerald, John and Wenger, Neil. *The Invalid Advance Directive*. Bioethics. V. 13 n. 2 pg. 32-34

Fodor, Jerry. "Where is My Mind?" London Review of Books, vol. 31 no. 3, February 12, 2009.

Follman, Mark. "Inside the Race to Stop the next Mass Shooter." *Mother Jones*. NNov./Dec. 2015. Web. 20 June 2016.

Firlik, Andrew D. "Margo's Logo." *JAMA: The Journal of the American Medical Association* 265.2 (1991): 201.

"Functionalism," by Thomas W. Polger, *The Internet Encyclopedia of Philosophy*, ISSN 2161-0002, http://www.iep.utm.edu/, June 7, 2017.

Gangopadhyay, Nivedita, Michael Madary, and Finn Spicer. *Perception, Action, and Consciousness: Sensorimotor Dynamics and Two Visual Systems*. New York: Oxford UP, USA, 2010. Print.

Gill, Michael B. A Moral Defense of Oregon's Physician-Assisted Suicide Law. Morality. Vol 10 n. 1 pg 53-67 (2005)

Grice, Paul. "Personal Identity." *Mind* 50 (1941): 330-50.

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

"History of Vaccines - A Vaccine History Project of The College of Physicians of Philadelphia." *History of Vaccines*. Web. 10 Oct. 2016.

Hobbes, Thomas. Leviathan.

Hochberg, Leigh R., Daniel Bacher, Beata Jarosiewicz, Nicolas Y. Masse, John D. Simeral, Joern Vogel, Sami Haddadin, Jie Liu, Sydney S. Cash, Patrick Van Der Smagt, and John P. Donoghue. "Reach and Grasp by People with Tetraplegia Using a Neurally Controlled Robotic Arm." *Nature* 485.7398 (2012): 372-75. Web.

Hölldobler, Bert and Wilson, Edward O. The Superorganism: Beauty, Elegance, and Stragngeness of Insect Societies. W. W. Norton & Company, 2009.

Hume, David. A Treatise of Human Nature. Oxford: Clarendon, 1983. Print.

IMDB.com. “Most Popular "Body Swap/Body Switching" Titles.” IMDb, IMDb.com, www.imdb.com/search/keyword?keywords=body-swap%2Cbody-switching&sort=moviemeter%2Casc&mode=detail&page=1&ref_=kw_ref_key.
Kant, Immanuel. The Metaphysics of Morals.

Kass, Leon. I Will Give no Deadly Drug: Why Doctors must not Kill. The Case Against Assisted Suicide. Ed. K. Foley and H. Hendin. Johns Hopkins Univeristy Press. 2002

Kean, Sam. The Audacious Plan to Save This Man’s Life by Transplanting His Head What: would happen if it actually works? theatlantic.com. September, 2016. accessed October, 12, 2016.

"King James Bible." *Official King James Bible Online: Authorized King James Version (KJV)*. Web. 14 Sept. 2016.

Levy, Neil. *Neuroethics*. Cambridge: Cambridge UP, 2007. Print.

Levy, Neil. "Rethinking Neuroethics in the Light of the Extended Mind Thesis." *The American Journal of Bioethics* 7.9 (2007): 3-11.

Locke, John, and P. H. Nidditch. An Essay concerning Human Understanding. Oxford: Clarendon, 1975. Print.

Lupyan, Gary. Swingley, Daniel. “Self-directed speech affects visual search performance”. The Quarterly Journal of Experimental Psychology, 2011

"Marlovian Theory of Shakespeare Authorship." *Wikipedia*. Wikimedia Foundation, Web. 20 July 2016.

Marquis, Don. “Harming the Dead.” Ethics. Vol. 96, pp. 159-61, October 1985.

Matrix, The. Dir. Lana Wachowski and Lilly Wachowski. Warner Home Video, 1999. DVD.

May, Thomas (1997). Reassessing the Reliability of Advance Directives. Cambridge Quarterly of Healthcare Ethics, Vol. 6, pp 325-338.

McClelland, Shearwood; Maxwell, Robert E. (2007). "Hemispherectomy for intractable epilepsy in adults: The first reported series". Annals of Neurology. 61 (4): 372–6. 2007

Miller, G. A. "The magical number seven, plus or minus two: Some limits on our capacity for processing information". Psychological Review 63 (2): 81–97. 1956

Milne, A.A. The World of Pooh: The Complete Winnie-the-pooh and The House on Pooh Corner. Dutton Children's Books. 1985

MotherboardTV. "Stelarc: The Man with Three Ears." *YouTube*. YouTube, 2013. Web. 09 Aug. 2016.

Mukherjee, Siddhartha. The Emperor of All Maladies: A Biography of Cancer. Scribner. 2010

Nagel, Thomas. "Brain Bisection and the Unity of Consciousness". Synthese. Vol. 22, pgs. 396-413. 1971

Nagel, Thomas. "Death". Mortal Questions. Cambridge University Press. pp 1-10. 1979

Nakic, Marina and Thomas, Paul. "Dissociative Identity Disorder in the Courtroom". Journal of the American Academy of Psychiatry and the Law Online January 2012, 40 (1) 146-148

Nelson, Jamie Lindemann [writing as James Lindemann Nelson]. Hippocrates Maze: Ethical Explorations of the Medical Labyrinth. Rowman & Littlefield. 2003.

Noë, Alva. Action in Perception. Cambridge, MA: MIT, 2004. Print.

Norris, Christopher. Philosophy Outside-In: A Critique of Academic Reason. Edinburgh University Press 2013.

Parfit, Derek. *Reasons and Persons*. Oxford: Clarendon, 1984. Print.

Partridge, Ernest. "Posthumous Interests and Posthumous Respect." Ethics. Vol. 91, pp. 243-64, January 1981.

Periyakoil, Vyjeyanthi S., Eric Neri, and Helena Kraemer. "Patient-Reported Barriers to High-Quality, End-of-Life Care: A Multiethnic, Multilingual, Mixed-Methods Study." *Journal of Palliative Medicine* 19.4 (2016): 373-79.

Periyakoil, Vyjeyanthi S. "Pitfalls for Proxies." Opinionator. New York Times. Web. 21 June 2016.

Perry, John. *Personal Identity*. Berkeley: U of California, 1975. Print.

Perry, John. "Personal Identity, Memory, and the Problem of Circularity." Perry, John. Personal Identity. Berkeley: University of California Press, 2008. 136-155.

Pitcher, George. "The Misfortunes of the Dead." American Philosophical Quarterly. 21.2 (1984): 143-64.

Plato, John M. Cooper, and D. S. Hutchinson. *Complete Works*. Indianapolis, IN: Hackett Pub., 1997. Print.

Polger, Thomas. *Functionalism*. http://www.iep.utm.edu/functism/. Accessed June 12, 2017.

Pulsifer, Margaret B., et al. "The Cognitive Outcomes of Hemispherectomy in 71 Children." Epilesia. Vol. 45:3, pp. 243-54, 2004

Rao, Jaya K., Anderson, Lynda A., Lin, Feng-Chang, and Laux, Jeffrey P. "Completion of Advance Directives among US Consumers." The American Journal of Preventive Medicine. Vol. 46. No. 1 pp. 65-70. 2014

Reid, Thomas. Essays on the Intellectual Powers of Man. Boston. Phillips, Sampson and Company. 1855.

Roach, Mary. *Stiff: The Curious Lives of Human Cadavers*. New York: W.W. Norton, 2003. Print.

Rovane, Carol A. *The Bounds of Agency: An Essay in Revisionary Metaphysics*. Princeton, NJ: Princeton UP, 1998. Print.

Ryle, Gilbert. *Concept of Mind*. 1949. Print.

Sagan, Carl. The Dragons of Eden. Random House, 1986.

Searle, John. *Minds Brains and Programs*. Behavioral and Brain Science. Vol. 3 no. 3 pgs. 417-457. 1980

Searle, John. Minds, Brains and Science. Harvard University Press. 1984

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

Serruya, Mijail D., Nicholas G. Hatsopoulos, Liam Paninski, Matthew R. Fellows, and John P. Donoghue. "Brain-machine Interface: Instant Neural Control of a Movement Signal." *Nature* 416.6877 (2002): 141-42. Web.

Singer, Peter. "Voluntary Euthanasia: A Utilitarian Perspective". Bioethics. Vol. 17:5-6, pp526-41, 2003

Singer, Peter. Practical Ethics. Cambridge University Press. 1976

Snider, Gordon. "Historical Perspective on Mechanical Ventilation: from Simple Life Support System to Ethical Dilemma." American Review of Respiratory Disease, 140(2_pt_2), pp. S2–S7 (1989)

Swaminathan, Nikhil. “Why Does the Brain Need So Much Power?” Scientific American, 29 Apr. 2008, www.scientificamerican.com/article/why-does-the-brain-need-s/.

Swindell, Jennifer Sproul. "Respecting Autonomy in Cases of Ambivalence regarding End of Life Decisions." Thesis. 2008. Print.

Taylor PJ, Kopelman. “Amnesia for Criminal Offences”. Psychological Medicine 14:581–8, 1984

TEDxTalks. "Alternate Anatomical Architectures | Stelarc | TEDxVienna." *YouTube*. YouTube, 2014. Web. 09 Aug. 2016.

Temple, Colin. "‘Just in Case’: If and Only If?" *ColinTemple.com*. 2012. Web. 20 July 2016.

Timmango. "Equal Pay for Monkeys." *YouTube.com*. YouTube, 2012. Web. 20 July 2016.

Tollefsen, Christopher. Response to “Advance Directives and Voluntary Slavery” Cambridge Quarterly of Healthcare Ethics, Vol. 8, pp 358-363. 1999.

Tomlinson, Tom. “The Irreversibility of Death: Reply to Cole.” Kennedy Institute of ethics Journal. Vol. 3 no. 2 pp 157-165. 1993

Tulving, Endel. “Episodic and Semantic Memory” in Organization of Memory. ed. E. Tulving and W. Donaldson. Academic Press. 1972.

Tulving, Endel. *Elements of Episodic Memory*. Oxford: Clarendon, 1983. Print.

Tulving, Endel. “Précis of *Elements of episodic memory*” The Behavioral and Brain Sciences. no.7 pgs 223-268. 1984.

Turing, A.M. *Computing Machinery and Intelligence*. Mind. (59). 1950. Pg 433-460.

TV Tropes. “"Freaky Friday" Flip.”, tvtropes.org/pmwiki/pmwiki.php/Main/FreakyFridayFlip.

Veatch, Robert M. "Doctor Does Not Know Best: Why in the New Century Physicians Must Stop Trying to Benefit Patients." *The Journal of Medicine and Philosophy* 25.6 (2000): 701-21.

Velasquez, Manuel G. *Business Ethics: Concepts and Cases*. Upper Saddle River, NJ: Pearson, 2012. Print.

Varelius, J. “Autonomy, Wellbeing, and the Case of the Refusing Patient” Medicine, Health Care and Philosophy, Vol. 9 pp 117-125. 2006

Verheijde, Joseph L, Rady, Mohamed Y, and McGregor, Joan. “Recovery of Transplantable Organs After Cardiac or Circulatory Death”. Philosophy, Ethics and Humanities in Medicine 2. Vol. 8. 2007

de Waal, Frans. "Moral Behavior in Animals." *TED*. Web. 02 June 2016.

Wade, Nicholas. "Scientist Finds the Beginnings of Morality in Primate Behavior." *The New York Times*. The New York Times, 2007. Web. 02 June 2016.

Welch, Ashley. "Russian man volunteers for first human head transplant." CBS News. CBS Interactive, 29 Aug. 2016. Web. 07 Aug. 2017.

Wicclair, Mark R and DeVita, Michael. "Oversight of Research Involving the Dead." Kennedy Institute of Ethics Journal 14:2 Jun 2004: 143-64

Wikipedia contributors. "Body swap." *Wikipedia, The Free Encyclopedia*. Wikipedia, The Free Encyclopedia, 7 Jan. 2018. Web. 25 Jan. 2018.

Wheeler, William Morton. *Ants; Their Structure, Development and Behavior*. New York: Columbia UP, 1910. Print.

Young, Emma "Infant rat heads grafted onto adults' thighs". New Scientist. December 2002.