University of Rhode Island
DigitalCommons@URI

Senior Honors Projects | Honors Program at the University of Rhode Island

5-2011

# What Is a Human Person? An Exploration & Critique of Contemporary Perspectives

Emmanuel Cumplido
ecumplido@my.uri.edu

Follow this and additional works at: http://digitalcommons.uri.edu/srhonorsprog

Part of the Epistemology Commons, Ethics and Political Philosophy Commons, Metaphysics Commons, Philosophy of Mind Commons, Philosophy of Science Commons, and the Social Psychology Commons

Recommended Citation
Cumplido, Emmanuel, "What Is a Human Person? An Exploration & Critique of Contemporary Perspectives" (2011). *Senior Honors Projects*. Paper 206.
http://digitalcommons.uri.edu/srhonorsprog/206

This Article is brought to you for free and open access by the Honors Program at the University of Rhode Island at DigitalCommons@URI. It has been accepted for inclusion in Senior Honors Projects by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons@etal.uri.edu.

# *What is a Human Person?*

by
Emmanuel Cumplido

HPR 401 Senior Thesis
16 May 2011
Professor D. Zeyl

Submitted in partial fulfillment of the requirements for the degree of Bachelor of Arts with Honors in Philosophy at The University of Rhode Island. The author hereby grants to The University of Rhode Island permission to reproduce and distribute publicly paper and electronic copies of this thesis and to grant others the right to do so. © 2011 Emmanuel Cumplido. All rights reserved.

# *Acknowledgements*

I want to express deep gratitude to Professor Donald Zeyl for sponsoring my Senior Project and sustaining an extraordinary level of patience as I sometimes too-slowly made my way through the research. I am thankful for his willingness to undertake the project with me as a learning experience and consider myself very fortunate to have been able to work with him during his last academic year as a professor of philosophy at the University of Rhode Island. I know that he will be missed and I hope that the end-result of our shared can add to the already sufficient list of goals he has helped students accomplish. I pray God blesses him as he moves on from the Professorship.

Thanks also go to several other professors at the University of Rhode Island for their conscious and unconscious contributions to my research during the past nine months. A large thank you goes to Professor William Krieger for allowing me to use an important book on Paul Churchland, an unconscious contribution as it was, and for presenting undergraduates with challenging material in the University's courses on Epistemology and the Philosophy of Science. Much of my research would not have developed as it had if it were not for those courses, neither would my interests have increased towards the philosophy science and epistemology. I thank Professor Cheryl Foster, Professor Zahra Meghani, Professor John Peterson, Professor Craig Nichols, and Professor Galen Johnson for teaching me in various courses. All learning is connected and my project has intersected with just about every philosophy course I've taken to date.

I give special thanks to Dr. William Young. His friendship, example, and advice in the past have been invaluable. I think it safe to say that if it were not for the many books I've read in his personal library, the challenging conversations I have had with him (on everything from Augustine to our independent study of Wittgenstein's Tractatus), I would not even have thought of this project. He has one of the most penetrating minds I have ever witnessed, and he's contributed more to my desire and ability to understand the world than any human person I've known.

Last but certainly not least, I thank my entire family for supporting me through my undergraduate career and being more understanding of my workload than most would be. I thank Adam, Bryant, Erik, Alfredo and Michael for being friends through my reclusiveness, Mr. and Mrs. Bankston for going beyond what's necessary to help me in innumerable ways for the last several months, and Bryna, for being my closest friend and helping me to both stay focused, and stay (somewhat) normal. Most of all, I thank God as my Savior, the ultimate source of all good things, and sustainer of my life.

# *Contents*

I. INTRODUCTION *4*

II. A METAPHYSICAL ARGUMENT AGAINST PHYSICALISM *6*

III. CAN A PURELY PHYSICAL THING HAVE BELIEFS? *12*

IV. INTENTIONALITY *17*
Hilary Putnam and "Twin Earth" *18*
Functionalism, Fred Dretske, and Moths *26*
Kim's Close and Dennett's Dodging *30*
Why Try Eliminativism? Stich and Churchland Answer *35*
Why Abandon Eliminativism? Boghossian on Content *43*
Microfeatures, Social Psychology, and Insensitive Seminary Students: Alternative Motivations for Eliminativism *45*

V. PHYSICALISM'S EPISTEMOLOGICAL PRECIPICE *53*
Preliminaries: Warrant and Your Brain on XX *53*
The Evolutionary Argument against Naturalism *58*

VI. PERSONAL IDENTITY *65*
What's a Ship? A Physicalist's Identity Crisis *65*
The Psychological-Continuity Criterion for Identity *67*
Organic Identity: Persons as Organisms and Processes *71*

VII. PHYSICALISM, DUALISM, & BIOETHICS *77*
Human Persons and Abortion *78*
Reproductive Technologies, Cloning, and End-of-Life Care *82*

VIII. SOME OBJECTIONS TO DUALISM *85*
What's the soul made up of? *85*
"Dualism is Anti-Scientific" *86*
Mind Kiss Matter? The Energy-Conservation & Interaction Objections *87*

IX. A DUALIST CONTRIBUTION TO QUANTUM MECHANICS *93*
Quantum Superpositions *93*
The Snow Leopard's Not Dead: Entanglement and Linear Dynamics *67*
Not So Super: The Measurement Problem *99*
Soul Scientists: The Dualist Interpretation *101*

X. CONCLUSION *104*

WORKS CITED *106*

# *I. Introduction*

The title of this thesis exposes its core question: "What is a human person?" This can be translated as a very personal, existential question for each one of us, that being: "*What* am I?" This question has been a subject of debate for millennia, and the answers that have garnered people's allegiance through history fall under two broad categories: "physicalism" and "dualism".

By "Physicalism" I mean the idea that everything about human persons, from our mental lives to our identity, is entirely determined by and dependent on the physical facts of the world, especially the physical facts of the human body. One of the earliest renditions of physicalism was the philosophy of the ancient Greek atomists. In their view, all of reality could be explained through two principles: atoms and empty space. As a consequence, people were thought to be nothing but assemblages of atoms in space; human persons are human bodies.

By "Dualism", I mean, at the least, a denial of physicalism. Not everything about us human beings is determined by physical facts of the world or our bodies. Plato's *Phaedo* presents one of the earliest philosophical endorsements of dualism by arguing for the existence of an immaterial mind, or soul, that is the *grounds for* a human person's identity and responsible for our unique mental abilities, such as logical thinking. The idea that a human person is, fundamentally, an immaterial mind or soul has also been a long-standing position for many of the world's major religions in both Western and Eastern traditions.

My position throughout this thesis will be that of a substance dualist. I maintain that human persons are more than just purely material entities. Human persons are to be thought of as things *distinct in kind* from purely physical objects. Most fundamentally, we are immaterial

minds, or souls. I use those terms interchangeably throughout because though their meaning differs, I will not be concerned with specifying the *kind* of substance dualism I think is best.[1]

What I will be concerned with in what follows is a critique of physicalism. With advances in cognitive science and a recent revival of academic interest in studying consciousness, the debate on human nature has been receiving some special treatment. Physicalism has emerged as the dominant perspective in academic circles and it has often been a presupposition in my undergraduate courses outside the field of philosophy. I will be presenting a few of the troubling consequences that physicalism has in relation to epistemology, personal identity, and ethics. To close, I will also give a brief apologia by responding to the most frequently cited objections to dualism, and point us to how dualism could even contribute to our ever-developing scientific understanding of the universe.

My conclusion will be that the problems facing physicalism are insuperable, and should move us into investigating what sort of substance dualism can resolve these problems.

1 This is because I have not settled that question for myself. My purpose here is to say that some kind of substance dualism is right, and there are many options, just as there are for physicalist theories

## *II. A Metaphysical Argument Against Physicalism*

Plantinga's Replacement Argument (RA hereafter) starts by presenting a possible situation where I exist when my body ("B" hereafter) does not. It leads to the conclusion that, because of Leibniz's law of identity and the law of non-contradiction, I am not identical to B, or any part of B, since I have the property "possibly exists when B does not" (Plantinga, *Against Materialism* 3). He paints this picture in the first-person, and I will follow the same strategy. The idea is that anyone else can go through the same steps, and come to the same conclusion and that therefore, no other human person is identical to their body either (Van Inwagen, *Plantinga's Replacement* 3).[2]

There are two presuppositions to the argument that Physicalist metaphysician Peter Van Inwagen points out, with which some physicalists would agree. There's the thesis that "Human persons…are substances" (1).[3] In addition, the definition of "B" must be neutral with respect to both physicalism and dualism. This is necessary if the argument is going to avoid being biased for either position. Van Inwagen proposes this definition, which I'll also assume:

> *"My body =df the living human organism such that it is possible for me to bring about changes in that organism without bringing about changes in any other organism (other than such organisms as it may have as proper parts)—and which is such that causing changes in it can cause changes in me and in no other person"*(5).

With these two preliminaries in place we can go on to state the argument.

---

2 For those who don't there will be other arguments presented for why other ontology's of human persons, or criteria for identity, are unsatisfactory in major ways.

3 Van Inwagen gives some characteristics constitutive of substances: "…they persist through time, retaining their identities while changing various of their accidental properties; they are not grammatical fictions; they are not "modes of substance"; they are not logical constructs on shorter-lived things (they are not entia successiva); they are not abstract objects (they are not, for example, things analogous to computer programs); they are not events or processes (Van Inwagen 3)."

There is a "macro" and "micro" version of the argument, and since Van Inwagen presents an argument against the macro version that he says applies equally to the micro version, I'll concentrate on the macro version. First, this picture occurs in a possible world that is not ours, with certain peculiarities that, strange as they are, are not impossible. At any given time in this world, one hemisphere of my brain is responsible for the totality of the processes and functions we usually suppose are done by the whole brain, including memory storage and recollection, and the other half is dormant, a literal space filler.[4]

Let's say then, that I am reading an article in the University of Rhode Island's student news paper, *The Good 5 cent Cigar*, about a peacock that escaped from the zoo. At midnight, in the middle of my reading the second panel, the following process occurs: Every part of my body, starting with my feet and continuing up with every regional part (legs, waist and torso, arms, neck), is replaced by a new part in succession (via any method you wish, Plato's Demiurge, some advanced alien medical technology), all the normal connections of the old part are re-established between the new parts and the rest of the body, all the way up to my brain. The way this occurs with my brain has specific parameters. The dormant hemisphere, call it H2, is replaced by a new hemisphere H2* and H2 is instantly annihilated. After this, the active hemisphere, H1, "transfers" or "copies" all the information to H2*[5]. Then, H1 is replaced by a new hemisphere H1* and annihilated the same way H2 was. This whole process, from toe to skullcap, takes one second, and throughout it I continue to read the second panel of the comic strip without noticing what has occurred at all (Plantinga, *Against Materialism* 4).

---

4 We know from real cases that it's not impossible for the brain to make serious adaptations in the location of functions necessary to live, we're stretching the notion for this exercise.

5 Plantinga thinks of it as a transfer of information, Van Inwagen gives the analogy of two boards, one has (x) switches in specific positions on and the rest off. The second has all switches off. Then the second is switched on into the same pattern as the first. Thus the information is "tokened" or "copied" onto the dormant hemisphere (5). I don't know which analogy is closest to the truth of what the brain would actually do.

Two things must be established for this to bring me to the conclusion that I continue to exist though my body does not. The first thing is that my body must really cease to exist during the replacement. I may think that, because at every moment during the replacement most of the parts of *a* body are in existence and connected to other parts in much the regular way, that my body never ceases to exist and that we're really just imagining a sped-up situation of something that happens all the time naturally (6). This is partially true; all your body parts are replaced over time. But the reason that B really does cease to exist is that there is an "assimilation time" for any new part of this replacement body, after it is in the appropriate spatial location and has the appropriate physical connections, to become a part of B. We can understand assimilation time with the example Van Inwagen uses of an eye being placed in an empty socket and being re-connected to all the appropriate nerve endings. This eye does not immediately become a part of my body; there is a causal process it must go through to start functioning in the body the way it is supposed to, ie all the relevant chemical processes happening in it (Van Inwagen, *Plantinga's Replacement* 9). Because this is a causal process, it will take time, and no matter how short the time (though it's definitely longer than one second) one could adjust the RA to make the envisioned replacement shorter than it (9).

So there will be a time, after all of B's parts have been replaced and annihilated, and during which *none* of the replacement parts have assimilated, that I will have no body. Plantinga rests his argument here, and indeed if I were to grant all of the above, most of which seems possible, then Plantinga's argument should compel me to believe that I am not identical to my body. However, Van Inwagen poses a relevant objection: "Why should I accept…that I should continue to exist throughout the…interval that contained the one-microsecond replacement episode" (10). The argument does seem to *assume* that I will continue to exist while this process

happens. Van Inwagen points out that the hidden argument for this belief is that "During the…interval, a single episode of conscious awareness occurs. If a single episode of conscious awareness occurs during a certain interval, a single person must be the subject of that episode. I am the subject of the earlier parts of this episode. Since a single person is the subject of the whole episode, I am therefore the subject of the final parts this episode" (11). But, according to Van Inwagen, what I should expect as I read the article during this replacement is for my consciousness to cease. I should expect the same results I would expect if I were "vaporized by the explosion of a hydrogen bomb", so far as my consciousness is concerned (11). Further, Van Inwagen thinks that in the scenario painted someone will come into existence who *believes* that they had read the peacock article to the climactic end of its being found in a chicken coop, but they would be deceived (11). They are in fact having *false* memories copied from my H1, and I am now dead, whatever dead means for Van Inwagen[6]

I respond with two things that I believe are relevant to Van Inwagen's objections and to anyone who would feel as though, because they have a commitment to physicalism, this argument cannot have any force and they should simply cross their arms in defiance. The first is that I do not believe that Van Inwagen's objection to how this argument proposes we find out whether a conscious thinker exists behind a continuous conscious experience is forceful.

Let's say that we were able to actually perform this replacement via some star-trek technology that, instead of converting my body to energy and then re-using it to form the same body, configures the exact state of my body, and then performs a replacement of the kind described above in rapid succession, annihilating my old body. Now this happens to me as I stand in a lab, three feet away from Van Inwagen (who in this world is a scientist!), debating

6 Considering his belief in the Christian doctrine of a future resurrection of all human persons who have ever lived.

with him whether metaphysician Stephen Hawking has been able to provide successful definitions of what an "individual material thing" is. Right after the replacement is done, Van Inwagen the scientist explains to me what has just happened and then claims that I am not the person who was talking to him three minutes ago. The reason he will give me for believing that I am not that person is that he believes I *could not have* continued to exist through the replacement, because I am necessarily my body. However, if I remember everything that has happened over the past three minutes, remember all the words of our conversation, and certainly do not recall any loss of consciousness, then where does that put Van Inwagen?

Van Inwagen objects to the idea of finding out whether a continuous conscious person exists during the replacement by asking the replacement person whether they were conscious the whole time, but I see no other way to find out besides merely assuming either physicalism or dualism (12). So, going back to my lab situation, Van Inwagen, and all physicalists, would have to hold the commitment to physicalism although the only relevant empirical data, my feeling that I have endured throughout the past three minutes, seems to falsify this commitment.

Obviously, this response is not meant to settle the argument, and a physicalist may not be moved by it. But I do think it's relevant, since many physicalists place so much importance on the results of experiments and relevant fields of science to provide support for their positions. It would be awkward for a physicalist to object to this argument because of skepticism about human technological capabilities in the future. If that were given up, then, in this hypothetical outcome of such an experiment, they'd find themselves at odds with the only source of data. Perhaps, if anything, this will soften up some physicalist's hardcore empiricism to see that their position rests on something more than just "evidence".

A more philosophical reason to why I think Van Inwagen's objection is not persuasive is that the replacement could occur while I am asleep and not dreaming. In that case, I would not appeal to my continued consciousness for the belief that I am still the same person that has existed throughout the experiment. I would simply appeal to my experience of being the same person. I would have absolutely no reason for doubting that I am in fact the same person who has existed for the past 21 years, as much as the physicalist scientist who invented this technology will try to tell me otherwise. In fact, I believe that if anyone were to put themselves in my shoes at that moment, they would see why the replacement argument offers a real door to believing that they are not identical to their body. The fact is, such a rapid replacement as the one I'm now talking about could have happened to all of us thousands of times in our lives and nothing at all would be different (except for the existence of some conspiratorial group of scientists responsible for these ongoing experiments, I suppose that's a big deal).

Plantinga calls this argument for dualism an "argument from possibility". I believe that after adding the two possible scenarios above, Van Inwagen's admission that he has "…not said anything that should convince anyone that either of these premises is false [the premise that a continuous conscious episode would occur during the replacement and that I, therefore, would continue to exist during the replacement] " becomes problematic for anyone who still wishes to maintain the belief that they are identical with their body (12). This argument sets down groundwork for the fact that there are more than just empirical considerations in the physicalist thesis and the arguments in the following sections for why human persons cannot be purely material entities are a force to be reckoned with.

## *III. Can a Purely Physical Thing Have Beliefs?*

The first epistemological objection to physicalism is that no purely physical thing could ever have a belief, and that attempts to make this seem plausible fail in one respect or another. Plantinga also develops an attack against the larger materialist enterprise based on this thesis, and I'll expound that argument for my purposes contra physicalism. If a physicalist thinks that there are such things as beliefs (some philosophers don't, and I will deal with this idea separately) then they will usually think of them as "neuronal events". They will have neurophysiologic properties[7] ('NP properties' from here on) and *content*. The content of a belief is "what it says about the world" (Brown)[8]. Plantinga explains that every belief has a proposition as its content. So, my belief that the Miami Heat will win the NBA playoffs in 2011 has as its content the proposition "the Miami Heat will win the NBA playoffs in 2011." Having those two terms at hand will help the rest of the way.

Here's a blunt way to put the problem to the physicalist: We can see that my belief that "the number 7 is prime" is about the number 7. But can, say, a tree trunk be about something? Can any artifact be *about* something? Can the atoms that make up an ancient city, in and of themselves, be *about* something? No.

However, there can be indication and indicator meaning, and physicalists all too often confuse these with belief and belief-content while trying to accommodate beliefs in their view on human nature (21). A tree trunk can indicate to a human that another human has been there before (say, by the clean cut resembling the work of a chainsaw), but there's no way for the tree

---

7 These are all the physical properties (neurons firing in a certain pattern in a certain place etc.) pertaining to the state of x's brain that are directly relevant to the belief "p" belonging to x., without which x would still be the completely the same except for the fact that they would no longer believe "p".

8 I assume that narrow mental content is the best way to conceive of this and I give an argument against Putnam's attempt to eliminate narrow mental content in the next section.

trunk to be *about* that human who once cut it. In the same way, our brains are made of physical particles, and there's no way for a neuronal event to *give rise to* or *be* the "aboutness" of our beliefs (21). The same goes for the artifacts and the ancient city. The ancient city *indicates* to an archeologist the presence of humans in the area at some previous time, and the archaeologist's belief that this is so has content that is *about* the city and that group of people. This is pretty straightforward.

Usually, an "indicator" is a natural sign of some other event because of causal or nomic connection (21). Smoke is caused by fire, thus it indicates that there is a fire is below the smokestack (21). The smell of hotdogs and burgers indicate a nearby barbecue, or town fair. Our bodies have internal indicators of blood pressure and saline levels. Obviously, none of these indicators are beliefs. So, how would a belief arise from indication?

Frogs have neuronal indicators of a fly when it flies by. We have a pattern of neuronal firing that goes on when we see, or are looking at, a tree. That tree being there, in part, causes the firing pattern (21). We can allow (for the moment) that such indicators could carry "content": "whatever it is that the structure indicates on that occasion" (22). This goes for our neural state when looking at a tree, its content is that there's a tree in front of us. We could also say that the indicators carry "information". However, being this loose should also allow us to say that my blood's saline-level indicators carry the content of whatever level of saline we have in our blood (22). This obviates it that indicator content has nothing to do with "<u>belief</u>, or belief content" (23).

In an attempt to hide this distinction, some philosophers talk about knowledge-gaining systems, or of evolution as an information gathering process, but no causal correlation entails belief. The information that is potentially in a system is not knowledge until it is believed. A thermometer does not believe anything about the temperature; whatever structure indicates the

saline content in my blood does not believe anything about it. Let us label the distinctions: the saline level indicators, or our neuronal indicators of the tree, can have *indicator content* but *not belief content* (23).

I must believe the content of my indicator for it to be a belief. There are times when I see something, but don't believe that my senses indicate to me what is really there. Say I see a Republican campaign poster for governor and the face on it is a splitting image of my friend Jim, but Jim's a libertarian. Though my senses indicate something, I don't believe their indicator content. Thus, it's not enough to explain belief content by simply equating it to indication content.

Someone may claim that a computer's memory can store the sentence "That humans are more than their bodies is an antiquated belief." That sentence has as its content the proposition "That humans are more than their bodies is an antiquated belief". So doesn't this mean that a material object can have propositional content? (24) Well, No.

These are cases of *derived* content. Sentences have derived content, the sounds we use are arbitrary and only meaningful by being put to use by humans, who already think and have primary content in their thoughts. The structure in the computer has *derived* content because *we* assign that configuration, and the resulting symbols, *that* content. Imagine if humans never existed. Now let's say a spontaneously assembled computer gathers a bunch of data through an artificial "eye" and consistently expresses the colors in front of its "eye" by words appearing on the screen. Does the computer *believe* the words appearing? No. It would be nothing more than a very sophisticated mirror, or interactive part, of the environment[9]. The computer will not *believe* or *think* that it is seeing red. It will just indicate that there's something red in front of it by having

9 I suppose I am denying such a view of humans, and inadvertently presenting another looming danger for the physicalist, the loss of personal agency.

the words "red" pop up on the screen. If God created the first human being on this planet and they discovered the correlation between the words on the screen and the colors the computer sees, they may come to have a belief with the content "when the letters *'r e d'* appear on the computer screen, there's something red in front of the computer.

All this being said, there is another objection that Peter Van Inwagen brings against this fairly powerful argument against physicalism. Van Inwagen thinks that dualism suffers from the same inability that plagues physicalists in trying to explain how a physical thing could think. How does a dualist explain how a non-material thing could think? He contends:

> It is just that it is a bit easier to see that thinking is a mystery when we suppose that the thing that does the thinking is physical, for we can form mental images of the operations of a physical thing and we can see that the physical interactions represented in these images—the only interactions that can be represented in these images—have no connection with thought or sensation, or none we are able to imagine, conceive or articulate. The only reason we do not readily find the notion of a non-physical thing that thinks equally mysterious is that we have no clear procedure for forming mental images of non-physical things. (Van Inwagen, *Metaphysics* 26)

That thinking is a mysterious activity is a given, whether or not one view makes it more mysterious than another is irrelevant to the truth of the matter. But to respond, I cannot form a mental image of many things that I still believe exist or are true. I cannot form a mental image of "the number's 79 being prime" but I *know* that the number 79 is prime. Even if we grant that we imagine faint "images" of numerals when we think of numbers, then having a mental image of something has nothing to do with possibility, because I can have a mental image of other things that are impossible, like the proposition "the number 79 is prime" being the color red[10]. The point is that one does not think it's impossible for a physical thing to have beliefs based on an inability

10 This is impossible because a proposition is content, and content can be translated, so my image is of the symbols.

to picture it. It's based on apprehension of impossibility, coupled with empirical evidence[11]. There is no equal apprehension for dualism. If there was, Plantinga points out a somewhat ridiculous result: "we'd have a quick and easy argument against the existence of God: no immaterial thing can think; if there were such a person as God, he would be both immaterial and a thinker; therefore…" (*Against Materialism* 30).

That's the offhand response to the physicalist. Van Inwagen also says that we cannot form any representation of how a physical or non-physical reality could undergird thought and sensation, so physicalism and dualism are again on par. This is not a problem for dualists. For the physicalist, thought is produced by the workings of parts, so a representation of how beliefs are produced makes sense. The dualist believes that persons do this by virtue of the *basic property* of being able to think, entertain propositions, and believe. And, lest someone say that this is a cop-out, basic properties are common to our understanding of the world and often go unquestioned. Electrons, for example, have the basic property of having a negative electric charge (32). Propositions have similar properties, two of them can be related to each other logically, and it's hard to explain *how* in physical terms, or explain it at all except by apprehending it is a brute fact. The same is true of the human person, they can think and believe because it is a part of their nature (32).

11 Apprehension that thoughts are about something, coupled with the fact that physical things are not about anything, for example.

## *III. Intentionality*

Picture the earth before any human walked its surface. Now, take anything there and, as we did in the previous section, ask the question "Is *x about* anything else? " Is a tree *about* the ground, or the animals living in it, or any other physical thing? No, a tree is not about anything, it simply exists and interacts with its environment according to the relevant physical laws. Now contrast that with the fact that when you are thinking and asking the question, "Is the tree *about* anything else?" you are having a thought *about* that imaginary tree. This "aboutness" is also called "intentionality". Now consider the fact that every thought you have is *intentional*, it is directed to something in the universe. At this moment, you are probably having a thought *about* this sentence, or a thought about the fact that your thoughts are *about* things. Thoughts are like that, they have a non-dispositional property of being about things. (Being a non-dispositional property just means that this property of intentionality cannot be further analyzed.)

If humans are wholly material beings, how is it that thoughts, which presumably are also material entities, have the property of intentionality? As Geoffrey Madell has stated "…it is quite unclear what an *analysis* of a thought's directedness to an object could possibly amount to" (*Mind* 11). Say we are looking at all the neurons in some person's brain as they look at a tree. Those events made up, ultimately, of subatomic particles are not *about* anything in the way that this person's thought "That tree is an oak." is about the tree in front of them. The tree itself is made up of the same kind of particles that this person's brain is. If we see the impossibility of any configuration in the tree to have the property of intentionality, we can similarly recognize the parallel difficulty in what it would mean for any physical state in the brain to be *about* anything else. It seems as though the non-dispositional property of *intentionality* that thoughts have is *irreducible* to any physical description, and thus any physical explanation.

In general, the three responses of physicalists to this are:

1. Despite the *almost* a priori difficulty, to say that these intentional states are in some broad sense identical to certain physical states and then to analyze these states into being dispositional (functionalist accounts are thought by some philosophers to fit into this category).

2. To admit that though these states are irreducible, this does not present any serious problems for materialism, and then to either analyze these states into being dispositional, or not do much of anything (versions of functionalism can fit here also).

3. To say that we should abandon all notions of intentional mental states (for example, stop thinking that there are such things as beliefs) as part of an antiquated view of human thought processes. These are "eliminativist" views. Many under this broad umbrella recognize the difficulty in reducing intentionality into anything else, but would simply state that concepts like "beliefs", which manifest irreducible intentionality, are illusory.

I have sectioned off my analyses of different philosophers' takes on this issue to make it easier for anyone who is interested in a particular philosopher I critique, and to simply make it easier for you to stop and come back to these arguments.

### ***Hilary Putnam and "Twin Earth"***

One landmark objection raised to the dualist idea of intrinsic intentionality has to do with what's called "the externality of the mind". For those familiar with the idea, I will not be dealing with that thesis as a whole, but rather only in its relation to intentionality. In *Reason, Truth and History* philosopher Hilary Putnam contends, "No set of mental events are sufficient, or even necessary, for understanding and reference" (quoted in Madell, *Mind* 15). What Putnam says could be grouped under response number 1 in the previous section. There is a direct implication

of this proposal for intentionality as we've construed it thus far. Mental events that are intentional (thoughts, beliefs, specific fears, etc.) are usually thought of as being used to understand and refer to the things in the world *as a direct result of* their being intentional. My belief "Snow leopards are the most majestic felines on earth" refers to snow leopards simply because beliefs have the ability to refer. We can refer because we have thoughts that are intrinsically intentional. We can vocalize some of these thoughts through sounds that our linguistic communities have agreed denote certain objects, in my case "snow leopards" for the most majestic feline on earth. But, if in fact these states are not used to refer, if, as Putnam says, they're not even necessary, it seems as though our notion of intentionality is placed on a fundamentally mistaken view of these mental states.

Putnam argues that the reason concepts refer to things is that there are *necessary* physical connections between a subject and the outside world, these physical connections are what the "intentionality" of thought can be broken down to mean, and without them reference is not possible. In section 6 of his chapter "Brains in a Vat" Putnam defends his idea that referring to something does not require *any* mental event, that the intentionality of a thought is not an intrinsic, un-analyzable, property of thoughts. I agree with George Maddell's assessment of Putnam's discussion when he says, "While it begins with a discussion of the notions of reference and meaning, towards the end of the chapter we find that what is being discussed is the notion of understanding" (Madell, *Mind* 18). I will treat section 6 of Putnam's chapter piece by piece.

He starts by saying that "If there are mental representations that necessarily refer to external things, they must be of the nature of *concepts* and not of the nature of images"(Putnam 396). Putnam clouds much of the issue by concentrating on images together with concepts. He states that if an alien sees an image of a tree, it's not a representation of a tree because there is no

causal interaction between the alien and real trees. I disagree. The *real* reason the image is not a representation of a tree is that the alien has not *thought of* the image *as* representing a tree (Madell 16). As Putnam says: "images do not necessarily refer", and that's the primary reason why the image of the tree doesn't refer to trees, *for the alien*. But if one could explain what a tree is to the alien, and tell them that the image they see is of a tree, then the image would represent trees, admittedly still imaginary things to the alien.

Moving forward though, Putnam claims that he could imagine someone "thinking just these words and having just the feeling of understanding, asserting, etc., that I do, and realizing a minute later (or on being awakened by a hypnotist) that he did not understand what had just passed through his mind at all, that he did not understand the language these words are in" (Putnam 397). Because this is supposedly possible, Putnam claims that therefore to say that someone is having a thought, or has a concept, is not to say they are having an introspective mental event, because the concepts that make up, or *are* thoughts, can occur in someone who has no idea what they mean. So *really,* concepts, and thoughts, are signs that do not intrinsically refer, they require something additional (causal connection between the thinker and referent).

But, the person Putnam asks us to envision is almost absurd, and surprisingly close to an example which would count as a mere "physical possibility", which he critiques elsewhere in his chapter. If someone really was "saying these words in his mind", and felt like they understood them, I have no idea what this "feeling of understanding" would be if not a thought *about* the words, because in my experience, feeling like I understand a word comes hand in hand with having a thought about the word, an intentional thought. Even if we grant that this person had a "feeling of understanding" and that they later realized they did not actually understand, they still had a thought *about* a bunch of words, *about* a bunch of arbitrary signs. That the words did not

refer to something because they did not understand them, is a different issue. Putnam says that this example shows that quote "the sign apart from its use is not the concept. And signs do not intrinsically refer"(397). If what Putnam means is that if someone says a word (a sign) in their mind, without knowing how to use the word, then they do not have the concept, nor do they understand the word, I agree. But, Putnam admits that private mental entities like thoughts are possible, and if he means that a thought *involving* a word (a sign), or a bunch of words, that I do not know how to use, is not intrinsically referential *because* I do not know how to use the words, then I disagree, because such a thought is still intentional, it's about the arbitrary words.

Putnam goes on to postulate the first of his "Twin Earth" examples to show how "meanings aren't in the head", and thus that the reference of a term in our thoughts is not determined by our mental states, and thus that referring (intentionality) is not an intrinsic quality of thoughts. I will quote Putnam's discussion and label parts of it for my analysis:

> "**[1]** Suppose you are like me and cannot tell an elm tree from a beech tree. **[2]** We still say that the reference of 'elm' in my speech is the same as the reference of 'elm' in anyone else's, viz. elm trees, and that the set of all beech trees is the extension of 'beech' (i.e. the set of things the word 'beech' is truly predicated of) both in your speech and my speech. **[3]** Is it really credible that the difference between what 'elm' refers to and what 'beech' refers to is brought about by a difference in our *concepts*? My concept of an elm tree is exactly the same as my concept of a beech tree (I blush to confess). (This shows that the determination of reference is social and not individual, by the way; you and I both defer to experts who *can* tell elms from beeches.) **[4]** If someone heroically attempts to maintain that the difference between the reference of 'elm' and the reference of 'beech' in *my* speech is explained by a difference in my psychological state, then let him imagine a Twin Earth where the words are switched…apart from the fact that 'elm' and 'beech' are interchanged, the reader can suppose Twin Earth is exactly like Earth. **[5]** Suppose I have a doppelganger on Twin Earth who is molecule for molecule identical

> with me... **[6]** If you are a dualist, then suppose my doppelganger thinks the same verbalized thoughts I do, has the same sense data, the same dispositions, etc. **[7]** It is absurd to think his psychological state is one bit different from mine: **[8]** yet his word 'elm' represents *beeches,* and my word 'elm' represents elms… **[9]** Contrary to a doctrine that has been with us since the seventeenth century, *meanings just aren't in the head*. (Putnam, 397)

Let's assume [1] (I in fact do not know what the difference between an elm and a beech tree is). We can agree with [2] also, as long as Putnam does not assume that he *intends* to refer to the same thing when he says the word "elm" and the word "beech". If he *intends* to refer to the same kind of tree when he says "elm" and "beech", then, *to him the referent of the words is the same*. It doesn't matter what the whole world thinks, when I say either "elm tree" or "beech tree" to him, he thinks I'm talking about the same kind of tree (this is going to be crucial). To [3] one should say, yes it is credible, and, no, our concept of elm trees is *not* the same as our concept of beech trees. At the very least (this is the case for myself as I write this), our concept of "elm tree" is separated from our concept of "beech tree" by the content "they are not beech trees".

Let's take an inventory of what I just said through some examples, before moving launching off to Twin Earth. Suppose Putnam is seeing an elm tree, and says "That's an elm tree". If he intends to refer to the tree he's seeing, then *we* say he is using the term correctly. Now suppose Putnam is seeing a birch tree and again says "That's an elm tree". If he intends to refer to the tree he's seeing, then *we* say he is using the term incorrectly. If in these examples Putnam does not *intend* to refer to the same kind of tree when he says "elm" and "birch" (If, as I said in reply to [3] he at least has the content "they are not birch trees" as a part of the concept "elm tree") then the meaning of "birch", he'll agree, is birches. He'll admit that he used the term incorrectly when someone points it out. However, both times, *his intent was a matter of his mental state and not altered by outside factors*. He *intended* to refer to the tree in front of him,

even if he was wrong about what kind of tree it was, and this intention was an intrinsic property of his mental state. Now it's time for take off.

Obviously, there's no harm in imagining [4] and [5]. I am a dualist, so I'll assume [6] is true. Here are the clinchers: If [7] is true, then it seems valid to assume that Putnam's doppelganger, like Putnam himself, cannot tell the difference between beeches and elms either. However, does Putnam know how to tell an elm tree from an oak tree? If so, then there's also additional sensory content to his concept of "elm", though he has no additional sensory content for his concept of a beech tree, and thus, when he sees them, he just thinks they are elms. These two additional considerations suffice to dismantle Putnam's argument.

Let's imagine the doppelganger is standing in front of an elm tree, like Putnam. If the doppelganger is in the same psychological state Putnam is in, as [7] says, then when the doppelganger says: "That's an elm tree", then, if Putnam can distinguish elms from oaks, the doppelganger is saying this sentence with the intention of referring to the tree in front of him, and he *intends* to refer to what everyone else on his planet would call "a beech tree"! This is because he has the same sensory content as Putnam does for his concept of "elms". He may not have yet learned that what we call "elms" are called "beeches" on Twin Earth, maybe because some human from earth visited and confused him. This means that [8] would only be true *for the doppelganger's community*. What the doppelganger intends to refer to is still is a matter of his mind. Now, if Putnam cannot tell the difference between elms and oaks, then there's *no content to the concepts except for the individuating principle*. This doesn't really matter either though. The community would disagree with the doppelganger's use of the term "elm", but what he *intends* it to refer to each time he uses it, is still a matter of his mind, even if it's just any tree.

What people around him *think* he's referring to, because of how *they* use the term, is a matter of their minds.

Let's say that in the example I just gave the doppelganger can tell the difference between beeches and oaks, in the same way that Putnam can tell the difference between elms and oaks, except that for the doppelganger the sensory contents he uses pertain to beeches. He *intended* to refer to what we call beeches when he (because of a lack of the sensory content that Putnam has about what we call elms) said (correctly in our opinion but falsely in the Twin Earth opinion), "That's an elm". This would mean there's a difference in his psychological state from Putnam's, and [7] is not true because Putnam's psychological state is that he intends to refer to a tree with the sensory content of elms. In this case [8] is true, but the reason why is still a matter of each person's psychological state.

That discussion was rather technical, but when understood it affirms that one's community and causal relations to an object affect how we describe, and how we determine what are, the objects of our thoughts. What Putnam has said thus far is arguing that what a concept means is not determined by what we intend, but that's only true if we define "meaning" without ever thinking about the subjective aspect of meaning! Thus far, the ideas that we do not have immediate and privileged knowledge of our thoughts, or that our thoughts do not have the non-dispositional intentionality that they seem to have, is not established at all.

Putnam goes on in the section to say that having a concept is not a matter of having images:

> A man may have all the images you please and still be completely at a loss when one says to him 'point to a tree'...He may even have the image of what he is supposed to do...acting in accordance with a picture is itself an ability one may or may not have...no matter what sort of inner phenomena we allow as possible *expressions* of thought...it is not the phenomena themselves that constitute understanding, but rather the ability of the thinker to...produce the right phenomena in the right circumstances" (398).

This is where things become very entangled. I will quote Putnam one last time and state why these examples do not help establish his conclusion either:

> …a man pretending to think in Japanese (and deceiving a Japanese telepath)...shows the futility of a phenomenological approach to the problem of *understanding*...On the other hand, consider the...possible man who does not have any 'interior monologue' at all. He speaks perfectly good English, and if asked what his opinions are... he will give them at length. But he never thinks...when he is not speaking out loud...[rather] he hears his own voice speaking...and has...a general 'feeling of understanding'...No one would hesitate to say that he was conscious...just because he did not think conscious thoughts except when speaking out loud. What follows...is that (a) no set of mental events-images or more 'abstract' mental happenings and qualities-*constitutes* understanding; and (*b*) no set of mental events is *necessary* for understanding. In particular, *concepts cannot be identical with mental objects of any kind*. For...we have just seen that whatever it (a concept) is, it may be absent in a man who does understand the appropriate word, and present in a man who does not have the concept at all (399).

Here's my critique. The first quote *only* establishes that if we want to say someone understands a concept, they should be able to produce the right kind of sentences and actions in response to, or accordance with, that concept. This does not mean that if we have a thought of the image of a tree without knowing what the concept "tree" is (like the man in Putnam's example), that our thought is not *about* that image. So, the thought's intentionality has not been broken down. The only thing that is established is how *we* as third-person observers, should judge whether someone else understands a concept.

The second quote, re-establishes the first quote's point with a very objectionable scenario. I would agree that if a man produces all the right behavior, and has conscious thoughts only when speaking out loud, that he can be said to understand. I agree because *he still has conscious intentional thoughts* about the concepts he is employing, even if only when he talks. If however,

even when speaking, he has no mental events corresponding to his verbal expressions, I think everyone should hesitate to say that he was conscious, or that he understood, because he could simply be a sophisticated machine.

My final thought, and really the whole point of my critique, is that Putnam's claim that "concepts cannot be identical with mental objects of any kind" only makes sense if we accept that 1) someone who has never had a thought could be said to "understand" anything, but his examples are not convincing, and 2) that someone could have a "feeling of understanding" a concept without having an intentional thought *about* that concept. The only reasons to accept either of these ideas seem prima facie, like the fact that they help to get rid of the irreducible intentionality of thoughts.

### *Functionalism, Fred Dretske, and Moths*

Fred Dretske, Jerry Fodor, and a host of others now operate with a view of the mind as computational. Thus, intentionality is to consist of being related to *physically realized* representations in one's mind, and thought processes are "transformations of mental representations" (Madell, *Mind* 26). However, this will not explain how a thought can be about something if the thought itself is a physical thing. Again, no physical thing can intrinsically be about another physical thing, it can only be about something *to* a thinker. If we say that this aboutness arises from the ability of a thought to be used in a certain way (dispositional analysis), whether internally or in behavior, my immediate self-knowledge of my own beliefs or other mental states is complicated by the fact that I will probably *not* manifest this disposition in many cases, but I am still "in" the mental state.[12]

12 Madell puts the last example more poignantly by reminding us that we often "entertain an idle thought (27)."

Madell keenly points out that "mental representations are the contemporary physical analogues of empiricist images (29)" He re-iterates "Hume's Problem": no manner of linkage between images will give you a thinking subject, or thought. This is as true for the idea of representations as it was for the "images" of old empiricism. Representations do not have any *aboutness* in and of themselves. Fodor and Dennett have expressed that tying these representations together via some "computational" or "inferential" process will dissolve the problem. This is not true. No matter how you construe it, the inferential process from one representation to another "cannot be from one *thought* to another" because a representation in itself is not a thought (29)! What this does is, again, make one of our most basic experiences, that of self knowledge, impossible, because I know what I am thinking of before any process of inference to other thoughts or beliefs is made (29).

But, let's not be hasty. One lengthy attempt to take representations and construct an acceptable account of intentionality, specifically of belief, is undertaken by Dretske. Dretske relies heavily on the idea of indication. Where Dretske and other physicalists often go wrong, is the jump from indicator content to belief content.

Dretske gives three criteria for a representation to be a belief:

1. The representation must be a part of a representational system that has the purpose of indicating what it does. This makes an indication a *representation*. A thermometer represents the surrounding temperature, though many other things about the environment could be calculated by the reading on the thermometer, none of which would be a representation because none are part of what the thermometer is meant to indicate (Plantinga, *Content* 19).

2. The structure C whose function it is to indicate x, has belief content *"only if C causes some motor output or movement* M, *and the explanation of C's causing M is C's carrying the information that it"* (19). Dretske uses the example of a moth that upon having certain brain processes indicate the presence of bats starts to "execute evasive maneuvers" (19). Dretske does not count the moth's indicators of the presence of bats as beliefs that bats are presence because it is not in virtue of their information that the moth is evading the bats, but rather it is simply a matter of mechanism, or "the moth's genes (20)." Thus the third criterion.
3. There must be some process of learning for the representation to be a belief. Typically, learning will also help to reduce possibilities of further action. Imagine a bird that gets to peck at a choice of three colored spots on a wall, which continually change positions, and every time it pecks the red spot it is rewarded with food. Eventually, it will only poke the red spot and the bird can then be said to *believe* or *proto-believe* that there's a red spot in front of it (20).

Now, for one, this account cannot accommodate necessary beliefs like 7+5=12. It is difficult to put this point more succinctly than Plantinga. "An indicator co varies with what it indicates; when it occurs, what it indicates also occurs (or probably occurs). *7+5's equaling 12*, however, always obtains; hence no-thing co varies with it; hence nothing indicates it (*Content and Natural Selection* 21)." This same problem arises when we talk about historical facts, which are "accidentally necessary" (Plantinga, *Materialism* 140).

The more serious objections are virtually the same however. First, criterion #2 is too strong. We can see that no motor output is necessary for me to have a belief. As Madell says "..if my

thinking of something is a matter of my being disposed to behave in some way, it must follow that I can have no idea what it is I am thinking of until this behavioral disposition is manifested…" (*Mind and Mat. 12*).

Most importantly, Dretske's criteria side step the more fundamental issue: how the bird (in criterion three) would gain the belief "the red spot will give me food", or whatever it is, and the intentionality of that belief, from the mental representations. Take the example of a thermometer that's designed to turn on a furnace when it goes below 67 degrees F. When it goes below this temperature it fulfills criteria 1 and 2. The only difference between this thermometer and the bird is the third criterion. But the third criterion sheds *no* light on the intentionality of the state being ascribed to the bird. Before the bird consistently pecks at the red spot it seems like we're supposed to think that the bird is mechanistically pecking at random spots on the wall, until, somehow, the fact that its body is being nourished when it pecks at the red spot transforms the representation of a red spot (which up until now has simply been part of a functional process causing motor output) into a belief *about* the red spot. I think that upon looking at the initial problem, this transformation is completely misleading. The representation was never a belief, neither could the fact that it is beneficial to the bird's health to peck at the red spot cause the representation of the red spot to all of a sudden attain the *intentionality* needed for it to become a *belief about* the red spot.

Another way of looking at this problem would be to picture the moth's indicators (in criterion two) as being *designed* (say by God, or humans via some future advancement in science) to indicate the presence of bats and cause the output of evasive maneuvers (Plantinga, *Materialism* 140). These indications are representations. The only thing missing from the moth is this process of learning. But we could imagine the moth trying several different patterns of

evasive maneuvers on several different occasions. Each time, it gets bumped by a bat and suffers just-so-slight disorientation. Finally, the moth tries a maneuver that prevents its being bumped. Henceforth, it employs these specific maneuvers when the bat-representations are present. Does this in any way *show* that *now* the neural representations for a bat being present are beliefs? It seems obvious that nothing with respect to the actual representations has changed; their nature is the same. The physical states still are not *about* anything. Dretske's claim that learning transforms these representations into beliefs is almost arbitrary.

Finally, we can re-construct the whole scenario by simply attributing beliefs to the bird the whole time, as a basic activity it engages in as a conscious, immaterial mind, and say simply that its wrong beliefs (that if it hits one of the spots it will go away, that if it hits the spots hard enough food will come, if it hits all the spots food will come) are experientially ruled out, not that it has no beliefs about the spots during this process and suddenly it possesses this singular belief that the red spot will give it food. Or, we could opt to not attribute beliefs to the bird at all. The central point of all this is that indicator content does not equal belief content, Dretske's dispositional account of representations cannot explain the intentionality of a thought (it could blatantly deny it), and only a subject already capable of thought can see a representation as *about* something.

### ***Kim's Close and Dennett's Dodging***

Moving forward, Jaegwon Kim, another renowned physicalist, in defense of some kind of functional account of intentionality says "…it seems to me inconceivable that a possible world exists that is an exact physical duplicate of this world but lacking wholly in intentionality (qtd. in Madell, *The Road* 1)." A dualist will agree with this if what Kim means is that a world duplicating ours physically will probably not lack intentionality (though I would not say it is inconceivable), because many features of our world only make sense if there are creatures with

immaterial minds capable of intentional thoughts in them. But if Kim means that we can picture a duplicate world first, remove the source of intentional states (immaterial minds), and then still expect there to be intentionality, the dualist would object. In principle though, there is agreement with what Kim has literally said, since many of the physical characteristics of the world are brought to be through intentionality on behalf of thinkers.

In his most recent work, *Physicalism or Something Near Enough*, Kim has argued that all of a person's mental life can be reduced, except for qualia. This is surely an over-statement. Kim thinks that we will *eventually* be given a satisfying materialistic account of things like belief, desire, and the intentionality of thought in general, but explicitly says "it is 'perhaps unlikely that we will have such definitions any time soon' "(qtd. in Madell, *The Road* 2). In case someone should miss this. Kim has just said there is no satisfying materialistic account of beliefs; none, zip, zero. Yet, the state of affairs in philosophy of mind seems to imply that this isn't a point physicalists have to deal with any longer. Many place faith in the "eventual" release of a satisfying account.

Here is the gist of another, similar, appeal by Kim in defense of the *potential for* a functionalist account: "Consider a population of creatures … that are functionally and behaviorally indistinguishable from us … If all this is the case, it would be incoherent to withhold states like belief, desire, knowledge, action and intention from these creatures" (qtd. in Madell, *The Road* 2).

Presumably Kim means to say that if there are a population of creatures that are functionally and behaviorally indistinguishable from us, *and do not have immaterial minds*, we should still attribute thought life to them. But, just judging what he says, a dualist can agree, because this is how a dualist would think that we generally go about judging whether someone

has these states anyway. Except, a dualist would also say that we do so on the basis of first-person awareness of our own intentional thought-life and resulting behavior (2). Madell correctly states that there would be two options for Kim to take with regards to what the functional re-definitions we await would do to intentionality. One is to say that these re-definitions eliminate the non-dispositional property of thought**,** and that comes with the plethora of problems I've addressed with respect to Dretske's analysis, and others I'll bring up with regards to eliminative materialism's stance on this issue; the other option is to say that the redefinitions will leave this feature as it is, an irreducible facet of human experience, in which case the materialist enterprise should be considered a failure.

Philosopher Daniel Dennett has also offered what he takes to be an analysis of the intentionality in our mental life via his notion of "the intentional stance." There are two aspects to this analysis. First, Dennett says that people ascribe intentional states, like "belief," "desire" etc. to others' behavior as a predictive strategy; when interpreting behavior this way, one is taking the intentional stance (Dennett, 15). This schema is extremely successful. It is in virtue of this predictive strategy that, in practice, we come to find out about the second aspect, which is that intentional states are *real,* because they pick out certain real patterns in the world. These patterns are not physical but rather "intellectual" patterns. He gives the example of a Martian looking on at the stock market and, having a *complete* knowledge of the physical forces at work, is able to predict everything that happens, but, because it does not take the intentional stance it misses "a real pattern", namely the real pattern of intentional behavior exhibited by the people involved in the market. There are three problems here, the first two of which are related, and Dennett, though acknowledging them, doesn't really meet the issues. The third, more severe problem, he completely ignores, like many physicalists have seemed to do.

The first problem is that a physicalist *should* simply go the route of saying that the Martian would see the intellectual patterns, but see them for what they really are: some pattern (even if broad) of neurophysiology, or elementary particles, and thus beliefs as we conceive of them would be eliminated. The second problem is that our prospects of ever discerning such a physical pattern are almost hopeless; for the dualist, and even some physicalists, this is completely impossible because there are no token physical events that indicate, say, anger, in themselves, it is simply by virtue of our own experience (i.e. our own intentional mental life) that we take certain actions as being expressive of anger. Further, a dualist would out rightly deny that any being having complete knowledge of the physical world, but no knowledge of human's intentional mental lives, *could* make consistently accurate predictions of the stock market.

Like I said, Dennett answers this challenge indirectly. In a later essay he describes a frog and *our characterization* of the frog's actions in terms of intentional stances, like the frog's "belief that you are behind him" or "desire to escape" when it jumps away from us, as useful to an extent, but not the full truth (Dennett, 109). Dennett claims eliminativism would be right in saying that this description *can* be reduced to neurophysiology, but he also claims that eliminativists are missing some kind of real thing when eliminating the reality of thoughts and these intentional states altogether. He stays away from ontological problems of his position by claiming that he thinks beliefs are as real as the equator, or the gravitational center of an object. He seems to think that because there is dispute about the ontological status of these things (are they physical realities?), but their existence does not threaten materialism, that there's an equal ambiguity with the ontological status of beliefs that does not threaten materialism.

Let's just deflate this. The reality of the equator is obvious in one form or another, even if only as a "real" conceptual tool. If it is physical, you can give a physical description of the

equator at any given moment, even if the atoms that make it up are not consistent[13]. The reality of thought is obvious in one form or another also. But are they, as a physicalist *should* believe, physical realities? We should not be satisfied by a philosopher's claim that "the equator's ontological status is mysterious" when there are ways to respond to the question and defenses of different sides available. Neither should we allow Dennett to simply write off beliefs as being "real" physical patterns but not explain how these patterns avoid complete reduction. This leads us to the third and fourth problems with Dennett's line of thought here.

Dennett does not give a satisfactory answer (any at all?) to the question of what it would mean for a physical thing to be a belief or for any physical set-up to produce the *intentionality* present in thoughts. Surely, my having a belief does not mean *necessarily* that someone uses this intentional strategy to interpret my behavior and ascribe to me the state of having a certain belief. That would sound like the idea of holistic interpretation, which, though true to an extent, overreaches its explanatory scope and makes my demonstrable experience of self-knowledge impossible. There are times I entertain a belief and no one has any idea, no one interprets my behavior, and I am not disposed to act in any way because of this belief (like my belief that if I had had the ability to type faster when I started this thesis I would have had much more written about Dennett's position by now. I'm not going to go take a typing class, and…)

Along these lines, Dennett simply writes off the reality of self-knowledge (in the sense of a private knowledge of one's own mental states) with a possibility, the possibility that we are conflating the intentional stance of propositions with some more basic neuro-physiological thing (which could either be defined reductively or functionally I suppose, either one of which has received no plausible account yet) which is *actually* belief (Dennett, 115). He returns to the

13 I personally think that the equator is better seen as a conceptual tool, a real idea, not a physical entity.

example of the frog by saying that it's useful to characterize the frog's behavior, as well as human behavior, as intentional, but that "In both cases behavior is controlled by a complex internal state " (115). The difference is that with humans, language helps to give an illusion of individuation for specific beliefs, through expressions, internal to the subject or made public, but this could simply be "an artifact of the environmental demand for a particular sort of act" (114).

This also can be dealt with pretty swiftly. Dennett cannot simply assert that the reason beliefs seem to be equivalent to propositions is that we do not yet know the complex internal states that actually produce them. The reason it's easy for him to do this is that he is addressing himself to other physicalists like Fodor who want to picture some sort of *"Mentalese"* language[14] that is behind our propositional attitudes in a broadly organized set of correlations. This more basic mental language, Dennett correctly points out, need not be arranged in the same way propositional content comes to be arranged. But this doesn't matter to the dualist objection, one that everyone should see for what it is. Should Dennett be asked to explain the fundamental objection, what any of these physically describable internal states that *are* beliefs and thoughts, along with their intrinsic intentionality, are themselves, or how intentionality could arise from such physical states, or how to explain the reality of self-knowledge, we can expect no answer.

### *Why Try Eliminativism?*
### *Stich and Churchland Answer*

Up to this point, I have argued that physicalists can give no account of intentional mental states without unacceptable consequences. There are some thinkers who have seen this difficulty and have taken what may be the most natural turn. Their move is to claim that all concepts associated with these mental states constitute a theory of human behavior that is completely

14 Which will have to be instantiated by, if not equivalent to, some token physical set-up, and how any physical state could give rise to a belief along with its intrinsic aboutness, or how such Mentalese statements could ever become a proposition, I still have no idea.

false. The common term to label this theory is "folk psychology". I will call it "commonsense psychology" (CP hereafter), since it's obvious that the term "folk" is meant to carry negative connotation. Anyway, the claim is that there is no need for reduction because these concepts, such as beliefs and desires, have no application to anything in the real world. What needs to be done is eliminate these concepts from our talk about humans.

There are many considerations that lead philosophers to think that we must eliminate intentional concepts. The first two that I mention can be found in the writings of Stephen Stich (Madell, *Mind* 64).

First, since science is concerned only with causally relevant features of the world, and the content of propositional attitudes is determined by invoking factors which are causally irrelevant (we saw this earlier in the case of how the linguistic community one is a part of may determine what one means by a term), it follows that in explaining behavior we should not invoke these attitudes, and further, that such intentional states do not exist (64).

The second consideration is tied to the first. It is that folk psychology wrongly describes intentional states as reasons for which humans behave as they do to (Taliaferro 73). An example drawn from social psychology involves a group of insomniacs who "are given placebo pills which they are told will have certain effects. When the effects happen as predicted…and the subjects are informed that the pills were in fact placebos, the subjects tend to invent reasons for their behavior that are more respectable" (73). The insomniac experiment is frequently cited, so it will be good to describe some of the details to explain the supposed problem.

One group of insomniacs was given placebos and told that the pill would cause a multitude of symptoms, all of which are normal symptoms for any insomniac; this was the "arousal group". The other group was told the placebos were supposed to facilitate sleep,

basically cause all the regular symptoms of insomnia to stop; this was the "relaxation group" (Horgan 206, 207). The researchers predicted that the placebos would cause particular effects described in attribution theory[15] to kick in and produce certain behavior. Specifically, the arousal group would fall asleep faster, writing off all their normal symptoms to being effects of the placebo, and the relaxation group would take longer to fall asleep, being worried about their normal symptoms not going away (Horgan 207). The predictions had overwhelming success. However, the subjects, when asked for what they thought were the reasons for their behavior, blatantly denied that the placebos had any affect and gave other reasons like "they usually found it easier to get to sleep later in the week". Thus, because many other experiments generate similar results, Stich argues that the mechanisms that actually produce behavior could be completely separate from those that produce linguistic behavior and the expression of propositional attitudes. (Horgan 206, 207). If that's the case then, in tandem with the charge that we interpret others' mental states using other causally irrelevant features, a central reason for maintaining the reality of CP, causal efficacy, is lost and CP seems like deadweight. Stich's examples, though interesting, do not lend as much weight his prescribed solution as may seem. A dualist need not maintain that humans are not subject to self-deception. As Taliaferro points out, the experiments can easily be accommodated into a folk-psychology model of explanation in that the subjects invented such reasons because they had "still other beliefs and desires about what they found embarrassing and appealing" (74).[16]

Taking Stich's first mentioned point in isolation, even if I invoke factors external to a person in characterizing their mental state, the fact that they have a mental state is a matter of

15"A theory of attribution is a theory of how ordinary people assign causes to events such as behaviors and mental states (understood broadly to include character traits) (Ravenscroft)."

16 A further response will be made below.

their experience. Stich is making a suggestion akin to the idea of holistic interpretation and Putnam's externalism, simply in different colors, and the point it's supposed to implicate still seems unwarranted. Though determining the content of a belief is influenced by the specifics of our linguistic communities, this does nothing to sway our security in the legitimacy of explaining many of a person's actions by their beliefs. Say someone believes "My lips are dry", and "This chap-stick will hydrate my lips", their subsequent behavior of applying the chap-stick to their lips could properly be said to be caused by those beliefs. By itself then, the consideration at hand gives little reason against the supposition that there are an abundance of situations where intentional states are causally efficacious, still less does it validate the eliminative move. This relates to a third consideration in favor of eliminativism. To see how, another example by Stitch will be helpful. Madell cites it as follows:

> Can we say, of the Russian soldier who suffered severe brain-damage in the Second World War which seriously impaired his conceptual network, that he believes that the pencil is made of wood, even though he can say nothing further about what wood is…Does the increasingly senile Mrs. T, who tells us every so often that 'President McKinley was assassinated', really believe this when she can tell us absolutely nothing else about the event? (*Mind* 65).

Because of such cases, it seems as though our interpreting of intentional states not only invokes causally irrelevant features, it's unreliable and rests on shaky ground. This translates into an eliminativist attitude by being combined with the factors just mentioned, like giving weight to the idea that intentional states are merely conventions employed by humans in interpreting behavior amongst each other, and nothing else. But, first, the case above is obviously not normal, our ability to ascribe beliefs is usually not so tenuous. Second, echoing an earlier response to

Dennett's intentional stance, the intended picture leaves us without a clue as to where such an interpretation of behavior would ever have gotten started. If "mental kinds have a purely nominal essence", and some individual did not first robustly experience and realize a belief as being a belief, before anyone interpreted it, and before seeing the action to be taken because of the belief, the whole enterprise of CP becomes a historical absurdity (66).

One more consideration for taking the eliminativist position is put forward by Dennett and takes its cue from B.F. Skinner's view that the postulating of "little homunculi" (what a dualist may say is an immaterial mind) to account for consciousness generally, and thus the intentional states in conscious experience, needs to be explained itself. So, Dennett maintains a pseudo-eliminativism in that though intentional concepts can be retained in psychology, we should strive to explain these homunculi by more, "dumber" homunculi, and continue breaking down the intentionality of what appears to be one subject until we have an explanation of the physical sciences (the homunculi become say, cells) (Taliaferro 35). One rejoinder must be kept in mind. How such an assembly of unintentional objects can constitute an intentional state is a phenomenon that needs explanation also, and why we ever started to categorize these "armies of idiots" as Dennett calls them, as "my belief that x", is absent from the entire conversation. Dennett is setting up a model for explanation that requires reduction to physical elements.

This is obviously not a serious threat, there will always be a point of explanation that cannot be "gotten past"; the charge of an electron provides a simple example from physics itself (Plantinga, *Materialism* 117). In the investigation of any natural phenomena, the push for a "deeper level" explanation will eventually come to an end. For our mental life, the end comes abruptly. Intentional states are non-dispositional and cannot be further analyzed to fit into any materialist explanation.

Taking up the more important considerations, we can now ask what the influential philosophers Paul and Patricia Churchland mean by their assertion that the intentional concepts (Churchland)we use in describing our mental life are nothing more than "a theory about human behavior". First, there is the relevant structure to which Paul Churchland points. Theories tend to make possible generalizations about the relevant domain. Take the following generalization from physics: for any body (y), there is a force which equals the mass of (y) multiplied by the acceleration of (y), or f=m x a. Churchland construes our ideas of beliefs and their resulting behavior in the same way: If x fears that p, x will desire that not p, He gives more complicated examples but one can see the point quickly (Churchland 71). Jose Bermudez sums up Churchland's picture of CP as "A particular conceptual framework for explaining social understanding and social coordination in which the propositional attitudes are central" (Bermudez 38). For now, we will keep taking it for granted that CP is a theory. States like beliefs, desires, etc., are simply explanatory tools. We'll sketch some of the reasons why Churchland and others think it must be scrapped in toto.

An over-arching charge against CP is that it hasn't progressed or afforded advances in understanding the things it is supposed to explain. Churchland states that our current understanding of how people act based on the propositional attitudes, is essentially unchanged from the time of Greek antiquity (Churchland 74). Further, there are a plethora of phenomena for which CP affords no explanation. Churchland cites "the nature and dynamics of mental illness, the faculty of creative imagination…the nature and psychological functions of sleep…the common ability to catch an outfield fly ball on the run…the miracle of memory…the nature of the learning process itself" (73). CP apparently gives us no way of understanding these processes

that (some more obviously than others) fall under the range of "mental". For any theory to fail on such a large scale deems it highly suspect.

Another broad charge is the fact that CP is not in sync with our maturing theories in the physical materialistic sciences. "In short, the greatest theoretical synthesis in the history of the human race is currently in our hands…But FP is no part of this growing synthesis" (75). CP does not seem reducible to any other field of scientific enquiry. Churchland even develops at length an analogy comparing the functionalist attempt to retain CP to an alchemist's attempt to retain "vital spirits" in the field of organic chemistry (81).[17] For Churchland, this charge combines with the earlier one to form a strong case that CP is completely false in its characterizations of human cognition, and should be discarded.

In brief response to these first two claims, CP is not supposed to explain the phenomenon of sleep and its benefits to mental health, or the catching of a fly ball, at least not the mechanics of it.[18] Also, there can be a case made for the progression of CP in our understanding of many other phenomena, pace Churchland's claim. Terence Horgan cites Neuropsychologist Richard Gregory's account of visual perception as "employing concepts recognizably like the folk-psychological concepts" (Horgan 200).

Another example of progression in CP returns to the experiments in which insomniacs are given placebo pills and their explanation as to the causes of their behaviors do not mesh with what the researchers deems to be the causes. The insomniacs perhaps were not aware of the beliefs that in fact caused their facility or lack of sleep, but this does not mean that their beliefs

---

17 Since functionalists are largely physicalists, I do not have a large stake in answering this attack. It only serves to move us into a position to assess what I think is the only consistent move for a physicalist, eliminativism.

18 I could say that the reason the outfielder is there in the first place, and is going to risk injury to catch the ball, is that he believes he can make it to the major leagues. A couple years down the road, when playing outside of the competitive arena, he may not dive towards the fence, because this belief is no longer undergirding his play.

were not efficacious in producing these effects. In fact, Stich characterizes the experiment as one in which such factors are at work: "attribution theory…predicts that subjects in the relaxation group s…will infer that their emotionally laden thoughts must be particularly disturbing to them. And this belief will upset them further.." (qtd. in Horgan 209). This now popular idea of "unconscious" beliefs and desires are dramatic updates in CP (207).

Anyhow, these are responses to a priori parameters on what CP is meant to explain, and, the "history of science is full of examples in which our pre-theoretical expectations…turned out to be quite misleading. For example, the demand…on optical theories that they account for facts…having to do with the physiology or psychology of vision" (200). Similar to Dennett's explanatory requirements, this strategy is not very persuasive. CP does find support in countless day-to-day activities. To point to a class of phenomena that it does not explain, but that no one expects it to explain, cannot be seriously expected to throw large doubts on the existence of the states postulated.

A few more objections before moving on; CP is used to identify causes, not form "new causal generalizations", though it indeed can do so. Historians explain almost exclusively by employing CP, it is ideographic as opposed to most sciences, which are nomothetic. That CP hasn't changed much is also expected if many parts of it are in fact correct. Further, to put it bluntly, arguments relying on the disdain of pre-modern-era concepts seem to be pretty gratuitous in the epistemic weight they give to factors which don't really imply their desired conclusions anyway.[19] Finally, to really render the second broad charge ("CP does not seem reducible…") inefficacious we can imagine that, if it is as serious of a matter as Churchland

19 like some unimpressive arguments against the existence of God. "People thousands of years ago believed in God, (fill in a bunch of irrelevant empirical facts here), obviously God is a vestige of an unsophisticated mode of human thought."

claims it is, it would in essence constitute an argument against the existence of God. God is a being whose properties no natural science could hope to reduce, therefore…[20].

### *Why Abandon Eliminativism?*
### *Boghossian on Content*

Philosopher Paul Boghossian, in his essay "The Status of Content", calls all the considerations for eliminativism that I responded to in the last section "toothless". He further contends that the arguments that are important for eliminativism share a dangerous feature. Though they are targeted at mental content, they make no reference to the "bearers" of content properties, just the properties, and thus they could be used to argue against the existence linguistic content in general (Boghossian 17). In other words, they are arguments against the idea of linguistic meaning, period.

He lists four kinds of arguments for eliminativism, which we have encountered in differing forms in the previous sections: arguments from the indeterminacy of content (apparently stemming from Quine but also referred to by Stich), the holistic character of content (owing to Davidson but employed by Stich also), the irreducibility of content (just spoken of by Churchland) and the "queerness" of content ("advocated recently by Kripke's Wittgenstein") (17).[21] To the benefit of content skeptics (in that it enables the arguments to apply in broader circumstances), and in sync with what may be the popular conception of content anyway, Boghossian assumes "that contents just are truth conditions." This just means that what a

---

20 I have taken up this method of informal reductio from several of Plantinga's essays. The relevance for this essay is seen more sharply when we take into account that sometimes explanations claim that God has acted in the world. Problems like conservation of energy and causation need to be addressed of course, and dualism must answer them also. This will be taken up later.

21 "advocated recently by Kripke's Wittgenstein" It also is similar to considerations we've seen: "…no real property could have the sorts of features that common sense considers constitutive of content". I suppose being put into the relevant relationships of propositional logic would be an example of such a "queer feature".

sentence says about the world, are its contents, and those contents determine whether the sentence is true (i.e. if what it says about the world obtains in the world, it is true).

So, these arguments implicate irrealism[22] towards truth-conditions.

There are different kinds of irrealism, but for eliminative materialism the relevant kind is an "error conception" of content. An error conception towards a region of discourse with predicates and statements involving those predicates, call the region of discourse "F", says that though the sentences in F are declarative, (i.e. the predicates denote possibly real properties), nothing actually has those properties, and thus all of F's assertions are false (4). Say the property P is a part of "F" (CP for our purposes), an error theory with respect to P can be summarized: " 'x is P' is always false"(12). Two things are implicated in an error thesis. First, "x is P" possesses truth conditions.[23] This obviously must be the case for the statement to be regarded as false. Secondly, this means that for all error theories there is a presupposition "…that the target sentences possess truth conditions (19)."

But, with the aforementioned conflation of content and truth conditions in mind, an error conception of content is claiming that: "…All sentences of the form 'S has truth condition p' are false, where S is to be understood as ranging over sentences in the language of thought, or neural structures, as well as over public-language sentences" (174). Plugging in an example relevant to CP would go something like this:

The following sentence is always false:

---

22 Boghossian uses this term throughout and gives a definition: "An Irrealist conception of a given region of discourse is the view that no real properties answer to the central predicates of the region in question" (2).

23 Boghossian's argument is flexible, what truth conditions are can be "on whatever construal of truth is favored." His essay describes robust and deflationary views, in the case of error theories, both are viable, neither salvages the theory.

" 'Mike believes that the horse is black' (S) has as its content (or as its the truth condition) 'the horse is black' (p)".

Returning to the general claim of the error conception of content, the problem is inescapable. The error conception of content "implies both that truth-condition-attributing sentences have truth conditions" (because "S has truth conditions p" cannot be false unless it has truth conditions) "and that they don't have them" (because if "for no S and for no p does S have truth condition p", then the sentence "S has truth conditions P" cannot have truth conditions). So, the error conception of content, the eliminativist's best way to formulate skepticism towards the reality propositional attitudes (and thus their intentionality) is self-contradictory.

### *Microfeatures, Social Psychology, and Insensitive Seminary Students: Alternative Motivations for Eliminativism*

Jose Bermudez recognizes the problem Boghossian poses for the eliminativist's usual attack on CP's propositional attitudes (to remind us, these are the intentional states of believing, desiring, fearing, etc., that x). If the arguments are employable against content in general, they result in a contradiction. Bermudez thinks that the method of arguing must take on different form, consisting in two steps. First is the now familiar strategy of reducing the explanatory scope of CP, thus weakening the popular idea that CP is really as useful as some deem it to be. But moving beyond this, the second step for Bermudez is to point out the "fundamental mismatch between…the model of representation implicated in…CP…and the family of models of representation that…provide the best general picture of how the brain can be representational" (Bermudez 36).[24] In reducing the explanatory scope of CP, Bermudez addresses two questions:

24 I argued earlier that what it means for a brain-state to be a linguistic representation, a belief, is not at all clear. Here we say a point of agreement for between the dualist and eliminativist. Instead however,

(1) Does successful social behavior always require explaining and/or predicting the behavior of other participants?

(2) In those cases where social behavior does depend on explaining and predicting the behavior of others, do such explanations and predictions have to involve propositional attitude psychology? (Bermudez 42)

To address (1) we are given several examples. He cites, from game-theorists, the prisoner's dilemma. Basically, the situation is that if prisoners A & B betray each other they each get five more years in prison. If one does, but the other doesn't, the one who did will be freed, the other will stay another ten years. If both keep silent, they will only stay two years. The "dominant strategy", the one that reaps the best results without considering the other's choice, is to betray. But if both prisoners kept silent, it would fare better for both of them (42). The moral of this whole picture is supposed to be that using the "dominant strategy" one can make a decision, and one doesn't really need to know what the other person will do. You don't have to try and predict their behavior, and thus you do not have to employ CP at all (43).

In on-going social interactions which are essentially "iterated" prisoner's dilemmas, one may, instead of trying to predict the behavior of the other, assess what they've done in the past and base your own actions on those:

The best known of these heuristic strategies is TIT-FOR TAT, which is composed of the following two rules:

4. Always co-operate in the first round
5. In any subsequent round do what your opponent did in the previous round.

(Bermudez, 44)

---

they're arguing for representations of an entirely different kind. But how this could make sense of our higher cognitive faculties seems less intelligible than the idea of reduction.

Again, the shift is from predicting the others behavior using CP, to deciding your own.

A more pedestrian example, shifting to situations better categorized under (2) earlier, would be our interactions with a waiter. Such social interactions may seem to require that I predict the waiter's behavior in terms of their beliefs or desires. But, contra this seemingly natural assumption, Bermudez submits:

> The social interaction takes care of itself once the social roles have been identified (and I've decided what I want to eat)…explanation and prediction need not require the attribution of propositional attitudes…We learn through experience that certain social cues are correlated with certain behavior patters on the part of others and certain expectations from those same individuals as to how we ourselves should behave. (45)

The weaknesses in these general arguments are easy to see. For the most part, we severely doubt that many real-life situations arise are clean-cut enough as these suggestions could be adapted to fit. Take the prisoner's dilemma. The prisoners obviously know each other by name at least. To be able to betray the other means that there has been some amount of shared experience. Are we to think that neither of the prisoners would even attempt to sketch what the other's character and beliefs are and base their decision at least in tandem with those considerations? I recognize that the model is just a model, but does it pose any serious threat to CP?

Or take the description of a familiar social interaction with a waitress. How do we understand what it is for another to "expect us to behave" a certain way? The waitress is expecting us to order food when he comes to the table, is not this essentially the same as them believing that we will order food? Would we have any real understanding of the situation if we eliminated our thinking of them as having intentional states? I am basically appealing to our

desire to explain. The fact "Waitresses take your order for food when you go to a restaurant." would not explain why anyone is a waitress in the first place, or why this restaurant's waitresses talk to you so much more politely, and without answer to that first question one would have to say that the whole situation would be rather absurd.[25]

Moving on towards Bermudez's second step and more novel suggestions, he points out first, that Paul Churchland sees cognition as resembling the processes in artificial neural networks (computational models of sorts) in which the brain's representations of the world correspond to vectors involving large populations of neurons that are "pushed through" matrices to other vectors (other populations of neurons) and the result is an innumerable possibility of representations and processes (46). This "seeing" of Churchland is more properly a future expected discovery of neuroscience. Supposing that cognitive science produces a system in which these multi-dimensional representations are governed by laws, correlating them to each other and to the organism's sensory and motor functions, the result would be that the components of the propositional attitudes (subjects, predicates etc.) could not correspond to these processes in any fashion because of the sheer complexity of such neural states and the lack of a satisfactory conceptual repertoire (48).[26]

If we identified groups of neurons that carry "specific semantic or representational contents…" (again, I deny that this notion even makes sense), then they probably do so in virtue of each individual neuron's specific activation levels, which represent microfeatures or

25 An answer to that question could be, roughly, "Waitresses wait as a means to earn money, or because they thoroughly enjoy the job, or both." This would yield implications for a waitresses' beliefs about what the result of their waiting on you will be, both monetary and/or emotional satisfaction for them.

26 It should be noted that physicalists themselves have been divided over whether this would actually be the case even if Churchland's expectations are fulfilled (Horgan and Woodward).

subsymbolic feature of the environment.[27] Bermudez goes on: "the crucial question is the relationship between this subsymbolic level of representation in terms of microfeatures and the symbolic level of representation in terms of object and properties (50)."

The relevant issue is to identify a level of neural representation that serves as a more fundamental tool than CP to navigate our social interactions (53). For the sake of not ignoring the substance of his argument we can take a look at some of the examples.

The first is tied to vision and movement. Bermudez describes the widely accepted idea of there being "two visual pathways " (53). There is "vision for action" where visual stimuli are projected to the "posterior parietal cortex", commonly inhibited in people with optic ataxia[28], and "vision for identification" where visual stimuli are projected to the "inferotemporal cortex" and is commonly damaged in people with agnosia[29]. One experiment records that though a subject may perceive two circles to be different in size, when they reach out for them, they reach as though they are the same size. Thus, the property of "graspability" seems to be outside of the domain of CP, and is not responsible for the way the subjects act, because the subject's belief that one circle is bigger than the other, does not correspond to the movements made (53). Other such examples identify sensory microfeatures that provide the explanation for certain behaviors that CP would misidentify, building a case for the mistaken application of CP to much of our behavior.

A category with seemingly more direct implications is "The Influence of Situation in Social Psychology". Factors like whether one has found a dime, ambient noise, whether one is

---

27 Perceptual content is not fully reducible to belief content, many of Bermudez's examples will relate to that. Our perceptions outstrip our conceptual repertoire, some contend, and in the same way our "multidimensional neural representations" are not fully captured by the content of propositional attitudes. But, eliminativism goes farther, and says that the content of the propositional attitudes are misrepresentations, not partial ones (49).

28 "Optic ataxia is characterized by an impaired visual control of the direction of arm reaching to a visual target, accompanied by defective hand orientation and grip formation" (oxford).

29 "Loss of the ability to interpret sensory stimuli, such as sounds or images" (American Heritage Dictionary)

running late, have been found to play drastically heavy roles in determining behavior, and altruistic behavior in particular (60). One experiment, amusing or disheartening depending on your perspective, involved seminary students going to give talks and passing a man laying out on a stairwell groaning. The only consistent factor in whether they helped was whether they were told that they were late (%10) or if they had time to get there (63%). The relevance of these experiments is that they cast doubt on whether CP's propositional attitudes really are the "springs of action" and whether situational microfeatures play a larger role, in-tandem with many other things like those perceptual microfeatures described above and others not explored by Bermudez.

One must wonder, when it comes to social psychology, whether in fact large umbrellas of propositional attitudes can be cast, irrelevant of the subject's "fundamental attribution error"[30]. These experiments should not surprise us much. If we believe that people are not generally selfless, these situations are in line with what we would cast as the CP beliefs undergirding all their interactions in the world. If we do not attribute robust altruistic character traits to many people, we will expect situational factors to heavily influence their decisions. If the seminary students believe they are going to be late, and they believe they could lose future opportunities if they stop, and they don't really believe that say, God will honor their helping of a man groaning on the stairs, they probably won't stop, because of those beliefs (and lack of a particular belief). How is this not an adequate explanation?

Bermudez admits that the examples pertaining to mechanical action earlier do not suffice to displace CP in the realm of social interaction, and I have intimated that situationist social psychology does not militate against the existence of intentional states playing some underlying

30 This is basically the fact that people will give "nicer" reasons for their actions where the data obviously imply other reasons. Stich's insomniacs are another case of this.

part to the results observed in their experiments. As it stands then, eliminative materialism has been stripped of the strength it has against CP that many have rhetorically given it.

Our survey of a number of different attempts to analyze the non-dispositional nature of intentional states, or propositional attitudes, or whatever the preferred phrase may be, into functional states, into mere tools for interpretation, or into non-existence, has presented a myriad of problems.

The functionalist enterprise is deficient in ways that a dualists and eliminativists draw attention to constantly, but which, in agreement with Churchland, are unjustly ignored. Reduction seems impossible and this is a problem. It should lead to an ontological distinction, or eliminativism, not an "anomalous monism" as Davidson suggests, whatever sense can be made of that. What are "mental" properties if not immaterial properties? How does a brain concentrate on an entity, like a belief, with immaterial properties? Intentionality is a Trojan horse, if the physicalist accepts its phenomenlogical characteristics, they must abandon their position.

Interpretative considerations either mistranslate our ability to self-deceive, or our inability to pin down the content of another's mental states, as cues to the unreality of intentional states. They also seem to make the unwarranted claim that "…the mental itself exists 'only as described'…" and leave us with a baffling problem of why any human in history would ever have started to use this interpretative scheme in the first place if it did not have some solid ground in their own conscious experience (Madell, *Mind* 77).

Eliminativism has prospects, and no one is going to stop the enterprise of cognitive science, but as it stands we seem more warranted in the belief that our intentional states have a robust existence, than that some future description of social interactions in terms of neuron-detected micro-features could seriously endanger that belief.

Lastly, let's remember that it's not just beliefs that are the problem, as Madell says, "merely entertaining a thought is just as impossible to accommodate within a materialist framework as a fully-fledged belief" (*Mind* 66).

# *IV. Physicalism's Epistemological Precipice*

Now, with or without eliminative materialism, I want to demonstrate that physicalism about human beings, conjoined with a naturalist worldview, leads any person to a state of skepticism that, at the very least, should provoke some re-consideration of their belief in physicalism. I will be using Plantinga's "Evolutionary Argument Against Naturalism" to show why this is the case. I take this to be a major consideration and, along with the two problems already pointed out, reason enough to reject physicalism.

## ***Preliminaries:***
## ***Warrant & Your Brain On XX***

The first preliminary is Plantinga's undergirding account of warranted belief. There are several requirements for a belief to be warranted. I'll state each and the reasons for their necessity.

The first requirement is that the belief be produced by properly functioning cognitive faculties (Plantinga, *Knowledge* 11). This is meant to modify the too-simple idea that, if generally reliable processes produce a belief, it's automatically warranted. Your sight may in general be reliable, but if you are intoxicated what you see will no longer be as reliable, and the beliefs formed during that period of malfunction (like, "Someone added more stairs to this staircase, it's never taken this long to get to the top.") are not warranted.

The second requirement is that the faculty at work has as its purpose the production of *true* belief (12). Say I have a deadly illness and come to believe, against all odds, that I will recover. The faculty at work may have as its purpose the production of what is probably a false belief, because it will have other good effects on me (i.e. a less painful and/or less difficult death emotionally or even heightening my slim chances at recovery). So, though this over-optimism is

actually the result of a properly functioning faculty, the purpose is not the production of true belief, thus is not warranted.

The third requirement is that the faculties producing the belief are doing so in the appropriate cognitive macro-environment. A cognitive macro-environment is the general environment in which our cognitive faculties were designed to function under. For example, the earth's atmosphere is the appropriate macro-environment for our visual faculties.

The fourth requirement is that the properly functioning faculty, aimed at truth, and in a favorable macro-environment, is *successful* in its attempts (12)[31]. We can allow for "success" here to simply mean that the faculty, when all the above conditions are met, produces significantly more than 50% true beliefs (pick any number your comfortable with, I would prefer something in the range of 80%). Imagine that a "junior deity", as Plantinga suggests, was responsible for the appearance of humans on earth. Unfortunately, its lack of maximal knowledge caused it to overlook some details and now, even when we are working the way we were designed to, in the right macro-environment, and are using the faculties aimed at producing true belief, we might still acquire mostly false beliefs because of the faulty design.

Finally, the last requirement is that the faculties be functioning in a favorable mini-environment. A favorable mini-environment is one in which the cognitive faculties producing the belief, which are successfully aimed at producing true belief when functioning properly, *actually do* produce true beliefs (Crisp 48). A mini-environment is, quoting Crisp, "a detailed state of affairs which includes all epistemically relevant circumstances obtaining when the belief issuing from E is formed" (43). A house of mirrors is a mini-environment, contained within the larger

31 The additional requirement of a favorable mini-environment in Crisp's essay actually makes it so that this final requirement is assumed, and mini-environments are a buffer to Gettier-style objections in specific circumstances. Either way, it's important for what follows and so I include it separately.

maxi-environment of earth. So, even if we are in a favorable cognitive macro-environment, like here on earth, we can be in an unfavorable mini-environment, like a house of mirrors. If we've never been in a house of mirrors before, the unfavorable mini-environment can deceive us into thinking there is a person to the right of us who is actually behind us, the most rational thing to do is to just trust your sight. In such a case, our sight would be functioning properly, and our macro-environment is favorable, but the micro-environment is misleading and some of the beliefs we form with respect to our vision do not have warrant for us[32].

So, the requirements for the warrant of a belief are that the belief be produced by properly functioning cognitive faculties who successfully achieve their purpose of producing *true* belief when in their favorable macro-environment and that the specific microenvironment in which the belief is formed is not misleading.

A second preliminary to the argument involves understanding the use of the words "rationality", and the idea of "defeaters" for beliefs. There are two kinds of rationality, and two kinds of defeaters that will come into play. First, there's "proper function rationality" (PFR hereafter) (Plantinga, *Naturalism* 205). Plantinga explains PFR so:

> "the rational thing to believe, in circumstances C, is what a properly functioning human being—more exactly, one whose cognitive or rational faculties are functioning properly in the relevant respects—would believe in those circumstances" (205).

32 There are important nuances to the epistemology, and I encourage a reading of Crisp's essay "Gettier and Plantinga's revised account of warrant" to see them. Though I have found some problems with Crisp's account myself and revised his definition of an unfavorable mini-environment in an essay. It is also important to note that environment means "states of affairs", not directly perceptible surroundings.

Tied to PFR, there are "proper function rationality defeaters" (PFRD hereafter). Proposition D is a PFRD for Proposition B for me at time *t* iff at *t* my "noetic structure"[33] includes B, I come to believe D at *t*, and any human whose cognitive faculties are working properly, who has the same noetic structure as I do, and who comes to believe D "but nothing else independent of or stronger than D, would withhold B (or believe it less strongly)" (208).

An illustration Plantinga gives here is if you ingested the drug XX, which you believe is designed to make the cognitive faculties of those who take it unreliable for a month, starting one hour after taking it, but unable to detect their own unreliability (208). In fact XX has succeeded in doing this for decades of clinical trials to 95% of subjects. You believe that you've taken the drug, but you keep believing your cognitive faculties are reliable (call this belief "R" hereafter) despite the improbability of this belief (5%) for one of two reasons, or both.

First, you cannot function in the world if you have serious doubts about your cognitive faculties (208). Second, your maintaining belief in R is part of the expected results of XX when it hits the bloodstream in properly functioning humans. If you detected the mentally confused state XX has caused, you'd be an evolutionary anomaly and probably would have some cognitive hardware that could in other scenarios be detrimental.[34]

One could tie PFR to our discussion of warrant by saying that PFR is the fulfillment of the first and third requirements, and whether or not the others are fulfilled is a matter to decide in the specific situation. In the case of taking XX, you would be rational in believing R, in the PFR sense, but would not have *warrant* as defined above because requirement two is not met (207).

---

33 One's noetic structure is the totality of beliefs one has, all of them being traced back to few, or simply significantly fewer, foundational beliefs. One such foundational belief we will look at in what follows is the belief that one's cognitive faculties are generally reliable.

34 This is just an observation on evolutionary adaptations. Seldom are mutations simply beneficial, there is often a give-and-take in play that would make the mutations detrimental outside of the particular environmental circumstances they arose in. Perhaps this immunity to XX is paired with dyscalculia.

The faculties working to keep R intact are geared towards either straightforward deception, or towards keeping a life-threatening skepticism at arm's length, they do not have as their purpose the production of true belief. There is a sense then, in which you *would be* irrational in keeping belief in R. This is what Plantinga calls the "alethic" sense of rationality.

Alethic rationality includes only the results of cognitive faculties whose purpose is the production of true belief. Returning to our example, you would have an *alethic* rationality defeater (ARD) for R. ARD's can be defined by adding this idealization to the definition of PFRD above: *All* of the cognitive faculties of the "properly functioning" human being placed in my circumstances have as their function the production of true belief and nothing else, and successfully carry out that function (209). So, there are alethic and non-alethic processes (over-optimism is an example of an effect of non-alethic processes), and the results of the two processes can sometimes conflict (like your belief that taking XX *should* make you unreliable, but still thinking that you should rely on your faculties, because otherwise, even moving at all becomes a serious problem).

Closing these preliminaries out, we can see that though belief in R an hour after ingesting XX is in line with PFR, belief in R does not have warrant and you would have an ARD for R. If it weren't for those other processes keeping you from rejecting R, you would give it up completely. Actually, Plantinga suggests that in this scenario you do still have a sort of PFRD for R. He calls it a "*Humean* rationality defeater" (211). This is because you can postulate what you think *another* person should believe in such a situation if *they* were only using alethic processes, and you "don't know of anything that distinguishes [your] case from theirs" (210). Thus, the proper functioning of your rational faculties results in the belief that R is probably not true for you.

The "Humean" part of this is that having a ARD to disbelieve R results in a state of skepticism which cannot be held for long *because* it conflicts with PFR, but which can be recognized in so much as we humans are able to "take a condescending and dismissive stance with respect to these promptings of nature" (211). In other words, you would be functioning ideally when you are able to recognize that R is unwarranted given your ingestion of XX, but you cannot maintain this doubt at all times because of conflicting pulls in PFR. Plantinga cites Hume's "Treatise" as a good expression of the frustration that can arise when ARD conflict with our PFR (Hume found that his own presuppositions logically led to a debilitating skepticism):

> Where am I, or what? From what causes do I derive my existence, and to what condition shall I return? …I am confounded with all these questions, and begin to fancy myself in the most deplorable condition imaginable, inviron'd with the deepest darkness, and utterly depriv'd of the use of every member and faculty (qtd. in Plantinga, *Reply* 210).

Nice as Hume makes us feel, we'll now talk about the argument with our definitions of warrant and defeaters in the background.

### *The Evolutionary Argument Against Naturalism*

"Philosophical Naturalism" is, basically, the belief that there are no supernatural entities (Plantinga, *Naturalism* i). The God of the Christian Bible, or anything like God, would be an example of a supernatural entity. Other examples would include Satan, angels, demons, and, of central importance here, immaterial souls. Thought of in this way, naturalism is close to materialism. But since differing definitions could cause problems, we can rely on the central notion being the first sentence of the paragraph and it will be sufficient. Naturalism entails physicalism with respect to human beings, so from here on out we are addressing naturalism (and

thus at least weakening the case for physicalism and opening the door for entities such immaterial minds to exist). Currently, the most accepted view of human origins is that given by the evolutionary story. Current evolutionary theory claims that all life has sprung from common unicellular ancestors by way of random genetic mutation being filtered through natural selection and genetic drift (2). To set the argument in its proper context, pace some popular but naïve views presented in many undergraduate university courses, evolutionary theory is not incompatible with supernaturalism. Many Christians, for example, believe that the evolutionary story told by the sciences is *for the most part* correct, minus metaphysical imports like naturalism, and that God in fact guided the course of evolution to produce organisms like our bodies, so that we could navigate the physical world in a sophisticated way.[35]

Putting that aside, we can take the evolutionary view of human origins and join it with naturalism to bring us to the consequences of the EAAN which I believe should make anyone with such views change their position.[36] Before doing that, I'll adopt the hypothetical situation that Plantinga does and imagine creatures that are like us in every respect except perhaps some aesthetic differences, like being two feet taller on average and having long necks. They have all the same cognitive faculties we have and have come to exist by way of evolution. We'll call them "Meses".

Now, evolution is "concerned" with behavior, with getting organisms from point A to point B so that they can eat, reproduce, and survive. Everything is meant to enhance those primary objectives. Patricia Churchland in talking about the brain puts it so: "Boiled down to the essentials, a nervous system enables the organism to succeed in the four F's: feeding, fleeing,

35 I couldn't argue for this here, but one could look to Francis Collins, Alvin Plantinga, B.B. Warfield decades ago, as examples of Christians who hold this as a possibility.

36 I suppose after some grappling with the argument, relevant literature, and their own intuitions.

feeding, and reproducing…Truth, whatever that is, definitely takes the hindmost (4).” Further, beliefs in general are by-products of evolution. If this is the case, then it seems very unlikely that such by-products were selected for on the basis of their truth. Actually, it seems like belief-producing mechanisms would be selected for only on the basis of their connection to fitness-enhancing behavior. The EAAN states that because of this situation, the probability that the Meses’ cognitive faculties are reliable belief-producing mechanisms[37] (“R” hereafter), given naturalism (“N”) and evolution (“E”) is low, or inscrutable (4). This is expressed as: “P(R/N&E) is low or inscrutable”.

Whether this is actually the case will depend on the relation between adaptive behavior, which is what natural selection is primarily geared towards producing, and belief (5). Plantinga lists “four mutually exclusive and jointly exhaustive possibilities” for this relation.

The first (P1) is straightforward epiphenomenalism. This would mean that beliefs have no causal efficacy whatsoever to produce behavior. Apparently this is actually a very popular view among biologists. If movements are determined wholly by biochemistry then beliefs are by-products, and not important to how the organism moves. One can easily see that in this scenario the Mesis should believe that P(R/N&E&epiphenomenalism) will be stupendously low, or inscrutable.

The next option (P2) is semantic epiphenomenalism. This means that beliefs can affect behavior through their syntactical properties[38] but not their semantic properties (6). For a physicalist, a belief must be a material thing of some sort, probably involving neurons. The syntactical properties of a belief then, would be things like “the number of neurons involved in the belief, the connections between them, their firing thresholds..” (7). But, one can quickly see

37 Again we can simply think of reliable as producing comfortably more than 50% true beliefs than false.
38 Basically, their material properties in the neural structures or processes.

why this will still leave the probability or P(R/N&E) as very low or inscrutable. Truth is a semantic property; beliefs are true or false in virtue of their meanings, their content, not by way of any neural structure or material properties. If semantic properties do not enter as one of the causes of behavior, truth cannot be selected for in evolutionary processes, and hence if the Meses' cognitive faculties consistently produced true belief, it would be quite a coincidence (a sort of naturalistic miracle).[39]

The third and fourth options (P3 & P4) are that beliefs are causally efficacious to behavior by way of their semantics and that they are either maladaptive anyway (P3), or adaptive (P4) (8). If they are maladaptive, obviously P(R/N&E) will stay low. If they are adaptive however, how does this affect the probability of R given N&E? Still not much it seems; just because the Meses' beliefs help it to successfully live, they need not be true. The crux of the matter lies in the fact that any one action can be brought about by a plethora of belief-desire combinations (8). They could eat fruits because they believe that some ingredient in fruits will keep evil spirits from inhabiting their body and they don't want evil spirits in their body. They could run away from a tiger because though they want to be eaten, they always believe there's a better tiger to be eaten by (8). On a more systemic level that would make sense to a naturalist, they could almost all have beliefs in things like a personal God, immaterial beings, and beliefs that their identity is grounded in their being some kind of immaterial entity themselves. From a naturalistic perspective all such beliefs would be false, but would serve some adaptive purpose. Since there are near limitless ways in which such systemic blunders could occur and yet produce successful behavior, the chances that the Mesis' cognitive faculties produce *true* belief more than

39 Here we can see a precursor to the argument as a whole. Imagine if the Mesis came to know about their condition and the semantic epiphenomenalism that pervades their belief-producing faculties. If they came to this information through their cognitive faculties (not sure how else they would) then this information itself would be suspect, and they would enter a sort of self-defeating nebulous of uncertainty.

50% of the time, and are thus reliable (a generously low request), are still extremely poor for them or inscrutable.

To put the Meses' problem more systematically I'll import Plantinga's use of the probability calculus[40] in his argument:

"P(R/N&E)= P(R/N&E&C) x P(C/N&E) + P(R/N&E&-C) x P(-C/N&E)"[41]

What's been done is to eliminate P3 from consideration since "its contribution to P(R/N&E) can safely be ignored"(9).[42] We then reduce P1 and P2 to their central thesis about beliefs: "-C", the idea that the content of a belief *does not* enter the causal chain leading to behavior. P4 is now "C" and means that the content of a belief *does* enter the causal chain of behavior and is adaptive. So the probability of R on N&E depends on how probable R is on N&E and C [*P(R/N&E&C) on the left*] or –C [*P(R/N&E&~C) on the right*], conjoined with the probability of either C or –C being true for the Meses given N&E (*the right side of each multiplication*).

Following Plantinga further, it seems as though "-C" is the natural view to take, for any Meses who is a naturalist. If beliefs are, most essentially, physical states, then content seems to be an unnecessary tack-on for the success of the organism. Finishing off this hypothetical scenario with Plantinga we can plug in estimates to the equation. If P(–C/N&E) is high, say .7, then P(C/N&E) will be .3; let's also say that "P(R/N&E&-C) is .2" This leads us to conclude that the probability of the whole equation, and thus of P(R/N&E) "will be at most .45, less than ½" (10)[43]. Because these numbers are being plugged in through nothing but intuition, one may

40 It seems to be an alteration of Baye's Theorem, I explain it's meaning.

41 "i.e., the probability of R on N&E is the weighted average of the probabilities of R on N&E&C and N&E&-C (weighted by the probabilities of C and –C on N&E) (10)."

42 This is just because there's no way to see how it would *raise* the probability of R.

43 Plugging in the estimates: P(R/N&E)= (1 x .3) + (.2 x .7)

think that a Meses should simply withhold belief about R, but now, the P(R/N&E) is, in the end, low or inscrutable. The end results can now be assessed.

All Meses who are naturalists have an *alethic rationality defeater* and a *Humean* rationality defeater for R, though they may have PFR in some sense. Like we said earlier, it may be that belief in R is in some way necessary for them to function without going insane, or to prevent them from becoming very depressed or paranoid about their every move. Nonetheless, this sense of rationality can be seen for what it is, a frail coping mechanism. The ARD cannot be rationally defeated. This is because any attempt to make R more likely will be done by modifying underlying beliefs or adding to their background beliefs or trying in some way to use the cognitive faculties that they now have a defeater for. Basically, the only way is to employ some sort of epistemically circular argument (242). They may still believe R but they cannot now *reasonably* (in the AR or ideal PFR sense) believe R. They must concede that their belief in R is the result of processes that are not geared towards true belief, and then stop thinking about it.

Another thing the naturalist Meses cannot now reasonably believe is, well, anything. This includes their belief in naturalism. Since naturalism is the reason that the probabilities within the original estimates are low or inscrutable, the Meses, if they care about being rational, should either give up belief in naturalism, or belief in R, or belief in reason. Obviously the whole deploying of the Meses as the species in question is only meant to lessen the blow of the argument and emotional kickbacks a reader may feel along the way. What we've been getting at is that any human who believes in naturalism and evolution cannot reasonably believe both. Rejecting naturalism opens the door for entities such as immaterial minds or souls to exist; it allows for physicalism with respect to human beings to be wrong. Exploring alternate

possibilities will, to the relief of those who make the move and investigate, offer solutions to the skeptical problems posed by this argument.

# *V. Personal Identity*

There is another fundamental aspect of experience that seems to pose a looming problem for the physicalist: personal identity through time. Most people live under the assumption that despite the plethora of changes in their bodies and mental lives (changes in character, beliefs, new and "lost" memories) they are fundamentally the same person as the one who was born on "their" birthday. This view of identity could best be identified as an "absolutist view of personal identity" (Moreland & Rae 179). What's being claimed is that personal identity conforms to Leibniz's law of the indiscernability of identicals: if x=y (if x is identical with y), then whatever is true of x is true of y and vice versa (56). Unlike a physical object, like a table, which could still "partially be the same" as an earlier table of which it has kept 50% of its parts[44], each person is who they are, could not be someone else, and cannot be only partially identical to themselves or other persons (179).

### *What's a ship?*
### *A Physicalist's Identity Crisis*

This natural conception of ourselves comes into conflict with a physicalist view of human persons in such a way that it should, again, provoke a re-examination of the view itself and of the epistemic weight being given to third-person descriptions as against the first-person descriptions of the experiences we have. A physicalist will have to either give up the absolutist conception of personal identity, or find reasons for defending it on physicalist grounds. With a popular example described by Peter Van Inwagen called: "Theseus' ship" we can explore these two options and bring out the problems facing physicalism.

---

44 One should note that this also means that the table is not strictly identical to the table it used to be, if its identity is only based on physical properties.

Theseus has a ship, call it the "Ariadne", and sails on it for 30 years, over the course of which each wooden plank and sail that makes up the ship is replaced. All the planks and sails are put in the same dump (Van Inwagen, *Metaphysics* 236). Then, Stilpo reconstructs a ship using the same planks and sails, and puts them in the exact same arrangement (say even with all the same nails) as Theseus' original construction. Now we have the original ship that Theseus constructed (O), the reconstructed ship that Stilpo has made (R), and the continuous ship which Theseus has been sailing on, the one *he* calls the *Ariadne (*C).

The problems and their relation to physicalism can be quickly seen. If a ship is just "a hunk of matter", then, if ship A is ship B they must have all the same components and properties (qua Leibniz's law). Obviously then, ship C is not the same as ship O (239). Which ship is the Ariadne? To Theseus, the answer to that question would be ship C, but this is not a telling fact about the nature of identity, it could just be a useful fiction of sorts, a way for the government to collect taxes. Has there been one ship that Theseus has been sailing on, or thousands of different ones simply named the same? It seems as though, if we apply Leibniz's law consistently, every plank replacement was the end of an old ship and the start of a new one. Is ship O the same as ship R? If (and only if) all the properties that were possessed by O are possessed by R, it seems as though we'd say yes.

These questions and answers are analogues of those a physicalist should ask about a human person. If human persons are identical with purely physical objects, namely human bodies, and a human body, like the *Ariadne*, changes all of its parts, which body is identical with the person? How could a person remain the same physical object through a complete replacement of all its parts? For the *Ariadne,* ship C was *not* the same as ship *O*. Applying this to yourself, a human person, this would mean, not that you have changed drastically from who you

were 15 years ago, but rather that there was no you 15 years ago; the person that was identified by your name 15 years ago was not you *at all*.

### *The Psychological-Continuity Criterion for Identity*

Because of the obvious problem in grounding personal identity in physical properties of composition, some physicalists adopt a "psychological-continuity criterion of identity" (PCC hereafter). This is the idea that human persons can persist through time, though *the bodies they are* do not, if the memories and mental properties of a body at time *t2* (your body at the present moment, say), arise through the right kind of causal process from those of a previous body at *t1* (say your body ten years ago) (239).

What would this process be? The task of delineating this without falling back, at some point, to commonsense notions seems daunting. Usually though, continuity of memories and beliefs are what make it obvious that the same person has existed through time, and this is the general idea behind this strategy. However, a science-fiction-like possibility that Peter Van Inwagen proposes is that PCC would allow us to transfer all the information from one brain, to another "artificially grown" and "blank-slate" adult brain (in an artificially grown, adult body), via some yet unknown technology. We'd then say that the latter (the previously blank-slate body) is now the same person as the person from which the information was taken, since it will now have all the same memories and beliefs as the first, perhaps even remembering that they were about to have a body transfer. This seems to imply that the person is not, strictly speaking, their body, but rather something else (memories and other information), and this in turn seems to mean that the PCC contradicts the basic physicalist premise that a person is identical to their body (239).

One immediate response by the physicalist could be that this particular method of transferring all the information would not be the right kind of process. An immediate response to that, is that if all the memories and belies are transferred, the burden is on the physicalist to answer why not. Is the process really what grounds identity, or the results of the process? For now, we will stick with memories and mental states being the relevant factor being sought by PCC. We'll address any thinkers who equate identity to a process below.

Is there refuge from the charge of contradiction facing the combination of PCC (thus the brain-info. transfer) and the belief that a person is identical to their body? Van Inwagen lists two options. One could hold to the relativity of identity, essentially meaning that "there is really no such thing as identity"[45], rather, there are a plethora of different kinds of "identity-relations" which we employ when we talk about different things (240). We say "is the same car as…", "is the same collection of atoms as…" and these denote slightly different relations with the word "is", though they all mean some kind of identity relation[46]. So, a physicalist may say that the "is" in the phrases "x is the same person as y" and "x is the same human organism as y" have different meanings, and thus being the same organism does not equate with being the same person (240)[47]. Still, when a physicalist who holds to the relativity of identity claims that "person x at t2 is the same person as y at t1" (again, even if they claim that it's different from claiming "x is the same human organism as y"), the relation denoted by "is" is probably that x's mental states have come to be from y's in a particular kind of way. In other words, they are still holding to a PCC for identity when talking about human persons, but do not face "formal contradiction" in so doing (240).

---

45 Or "there is no one relation properly called 'identity' full-stop"

46 The distinction here is not between the "is of identity" and the "is of predication", rather it is between different "kinds" of "is's of identity".

47 How to do this I know not, and thinking it through some will make it obvious why few philosophers support the view.

We'll import another thought experiment to shed more light on how the PCC fails to achieve its purpose. Since thus far, even if someone ascribes to the relativity of identity, the PCC seems to be the only way to ground personal identity. Say that at some point in the scientific future a scientist is able to split the left and right hemispheres of your brain along with the left and right parts of your body (as much as could be done) and reintegrates each with a separate half-body without a brain (Moreland and Rae, 184). To complicate this a bit further, both of the resulting persons have the same memories and character as you did before the surgery[48]. Which of the two persons are you now? You cannot be both in the same sense[49], and drawing on the relativity of identity to postulate two different identity relations between the original body and the two resulting indistinguishable persons doesn't tell us anything about the kind of relation that makes one of the bodies you, since the experiment occurred in virtually the exact same manner for both resulting bodies. Neither of the two responses available seems to rescue PCC here. If you have completely ceased to exist, then continuity of memories, mental states[50] and character traits is not sufficient for grounding of identity, because both resulting bodies have those[51]. If only one of the resulting bodies is you, then a physicalist, up to this point, is at a loss to explain what about the body makes this the case, since both persons in the half-body transfer came to be through the same process.

The other option Van Inwagen lays out for responding to the brain-info. transfer scenario we mentioned at first, and now also the half brain/half body transfer, is that a physicalist could

---

48 Not preposterously far fetched, brains can take over most of the functions of damaged hemispheres.

49 Insert anything else as an example here to demonstrate the absurdity: "I cut the table in half and connected each half to two other separate halves. I guess that now, both tables are identical with the first one."; no, one is, or neither is, what's certain is that there are now two distinct tables.

50 There is a sense in which dualists believe that any given mental state is necessarily owned by whoever has it and that *that* aspect of mental states cannot be transferred (Moreland and Rae 160).

51 Perhaps, again, they have not come to have them by the appropriate kind of causal process, but this is grounding identity more in the *process* and not in the memories, mental states, and character traits. This idea will be brought up below.

maintain a four-dimensionalist view of identity. This basically means that objects are not merely three-dimensional, but four-dimensional objects *extended* in space-time (Van Inwagen, *Metaphysics* 241). As such, they exist in a "region of space-time that extends…from the first moment of...[their] existence, through the present, to the last moment of… [their] existence—if there is a last moment of…[their] existence…(241)" Going all the way back to Theseus' ship, this theory would maintain that ship (C), which Theseus has been riding on, *is the Ariadne*, and all the individual planks have parts (like the part extended from June 1700-July 1702 for plank #28) which are parts of the ship *Ariadne*. But, no plank is ever part of the *Ariadne, full-stop*; only the four-dimensional parts during which the plank contributed (or is contributing) to the structure of the *Ariadne* are.

So, though at any point during your four-dimensional existence, you are made-up of three dimensional atoms, no atom is a part of you *full-stop*. All the atoms which compose your body can change, and there is no contradiction in saying that you have continued to be the same person, your identity lies in the full, four-dimensional aspect. But, it does seem like one needs to add a PCC, or something like it, to tie together all the "three-dimensional cross sections" of your four-dimensional existence and make it so that only one specific body is *your* body. With the *Ariadne*, though four-dimensionalism justifies the idea that ship C *can* be the same ship, there's no positive reason in four-dimensionalism by itself for thinking that the ship Theseus has been on for the past 30 years has been *one* ship besides Theseus' labeling it so[52] (242). In adding the PCC here to ground identity however, four-dimensionalism opens itself up to the same kind of problem already raised to PCC, and is really no solution.

52 In studying identity more, I am drawing closer and closer to the idea that identity is fundamentally reliant upon the mental, my other essay on Van Inwagen's chapter on "Individuality" in his book "Metaphysics" brings forth how establishing the individual identity of any physical object is problematic. Though I did not write this in that essay, I think that appeal to the primacy of the mental is the way to resolve the problem.

For example, how do we apply four-dimensionalism to our two sci-fi surgeries? Perhaps four dimensionalism would say that brain-information transfer does result in the same person. But, on four-dimensionalism, which person is you (if any) after the half-body transfer? It seems as though the stalemate is still present for four-dimensionalism since one needs a PCC to separate off human persons as distinct individual entities in the first place, and both persons in our scenario come to be through the same general process (thus we have no reason to think they're different persons from the original person who was "split", but that can't be right).

### *Organic Identity*
### *Persons as Organisms and Processes*

In any event, Van Inwagen points out that both the relativity of identity and four-dimensionalism are unpleasant and carry results that may not be worth the price of maintaining physicalism.[53] His proposal is to instead make it part of the definition of an organism that it can change its parts over time (245). He quotes physiologist J.Z. Young in saying:

> "*The essence of a living thing is that it consists of atoms of the ordinary chemical elements we have listed, caught up into the living system and made part of it* ***for a while****"*(244) (bold emphasis mine).

But, with this idea in place, one can now take a puzzle John Foster discusses and pose it to Van Inwagen, Van Inwagen being a physicalist who does not deny the existence of the mental:

> *"Someone treads on Jones' foot…Our normal practice is to ascribe the pain-state to Jones himself…But why not ascribe it instead to some corporeal part of Jones* [like his

53 He quotes Judith Thomson's essay "Parthood and Identity across Time" in saying that four-dimensionalism is "a crazy metaphysic", and that "Very few philosophers have any sympathy with the theory of the relativity of identity" (242). This is not a refutation of course. I just mention that because to treat the whole view on its own would be a job I'm not able to undertake just now.

> foot]...*Or alternatively…to Jones together with the room…or even to the whole physical world" (Taliaferro 167)?*

The point here is that if there are mental states characterized as "being in pain", whether supervening-on or identical-to a purely physical state (or states), there seems to be, for the physicalist, no necessary distinction as to what part of the physical cosmos *has* that experience as its own (167).[54]

Van Inwagen (or any physicalist) could say: there are necessary "law governed correlations between the brain and mental life..." which make it such that only "certain sorts of biological animals can be persons", and as such only those animals can be said to have experiences (something akin to the PCC). But, then it would seem to follow that the brain, or some other part of the body, is what "has the experience". Surely our nose does not experience the pain when our foot is stepped on. The implication is that the subject to whom we attribute experiences could be identified with a body part, in this case the brain. This would mean that the criterion of identity that Van Inwagen has proposed will fall apart. We're no longer talking about an organism as a whole that can change parts, just a part of the organism that is at the metaphorical heart of their identity. But now the physicalist would be open to the criticisms already mentioned, like how can a brain be identified as the person if *all* of its parts have changed over time? We'd have to add "being able to change all its parts" to the definition of a brain, and the same question arises to what grounds the identity of the brain as the same brain and what part of it is the "subject" of a pain-state. The physicalist cannot answer Foster's

54 This delves a bit into the idea of whether subjective experience should hold epistemic weight. If we think that the phenomena of "having an experience as one distinct conscious subject" is something to pay attention to, then *who* this subject is will be answered by establishing what it means for a person to have identity through time, and I maintain that dualism does a better job of this.

challenge of identifying the individual person that is the subject of experience if he maintains that the identity of a person is grounded in their being a body[55].

However, if a physicalist still wishes to maintain, without further argument, that it was Jones as a whole biological entity that had the experience it seems as though they would be identifying Jones with a token of a type of physical process. What would "Jones was stepped on and is in pain" *mean* but a particular assemblage of matter being stepped on, nerves sending electrical impulses off in a *particular* way up the central nervous system through the brain, then ending at some point perhaps after causing reactions in other parts of the organism, none of the parts being *conscious* subjects[56] but merely conduits of physical processes?[57]

To say "Persons are processes" does not just mean that persons undergo, sometimes drastic, changes in physical appearance and personality. Physicalists and dualists alike would agree to that. Neither does the idea have anything to do with things like coming of age or learning. Think of other physical processes: chlorophyll production in plants, metamorphosis in insects, the water cycle from above to below the ground and back again, oxidation of metals. The problem for the physicalist is that there is now *no* unified person, only an abstract process like those just mentioned which we use to label certain organisms (perhaps "a human life"). Equating

---

55 A possible answer for the physicalist is to say that part of a person can have an experience, but that doesn't mean that that part *is* that person. I find this hard to believe though. Think of someone who loses as many body parts as possible without physically dying. Have they lost parts of their "self" (not "themselves")? If so, have they lost their identity (not, do they feel different or lost part of what they felt made them "individuals", but are they different entities)? If not, then it seems like the physicalist is still leaning on either the brain, or a PCC, to ground personal identity.

56 None of my neurons are in themselves thinking about how heavy this man is, and as I expressed earlier in this thesis, it's unclear how a billion un-thinking neurons can generate a unified conscious subject.

57 In a simple resolution to all this, dualism says that it was Jones as a basic, nonphysical subject that was in pain by virtue of experiencing the physical state he was in.

a human person with a physical process is essentially an eliminative view[58] that destroys the conception we have of ourselves and denies that there are any such things as lasting persons.

Tying back to Foster's challenge regarding the ownership of mental states on physicalism, we have a basic experience of psychological unity which grounds our experience that physicalism cannot account for. We can express this so: At time 1 you experience x, y, z. (say the sight, sound, and smell of a pot of soup as you walk towards it). At time 2, you experience x2, y2, and z2 (sights, sounds, and smells from the same pot but from a closer perspective). At time 3, you experience x3, y3, and z3 (same idea, now standing in front of the stove-top). At each time you are aware that only *one* person is experiencing the x's, y's, and z's, and that it is the same person each time, you (Moreland and Rae 183).

Reasoning through the premises of an argument also requires that one person be present through all the thinking. If there is no enduring self, a logical thinking-through of an argument by anyone seems implausible. Physicalists can respond "as each person-stage emerges or ceases to be in a series of 'thinkings,' that stage passes on to the next one its content and a feeling of ownership ('this thought was mine')"(187). As Moreland and Rae point out, at this juncture all that can be done is to offer the total picture as reasons for someone to seriously doubt that this makes sense of their experience, and we are getting close to the heart of the matter (187).

Hume raised an objection to absolutism (and dualism specifically) relevant here by claiming that he was never aware of himself, but only of sensations. Hume would say you were aware of the x's, y's, and z's above but not of yourself, because there is no sensation of a perceiver behind the senses. The problem with this objection is somewhat obvious. How would

58 I am not equating the eliminativism of say, Paul Churchland, to being an eliminativist view of persons. Though it's obvious to many that eliminativism with respect to the mental implies that our conception of ourselves will be shattered and will result in an elimination of identity.

Hume, or any physicalist, know where to look for the awareness of self to know that it's missing? What does it even mean to have looked for this sensation, or to not be able to find it, if there is no substantial self undergoing the looking, or disappointment at not finding it? Further, it's obvious that there will be no sensation analogous to sight or smell when being aware of yourself, but there is a mode of awareness given to us that seems pretty basic: while Hume was looking for the sensation of the self, there was one person doing all the searching and having all the sensations as their own, that he would miss this as a signpost to the underlying reality is more the result of his metaphysical commitments than lack of any support (190).

Such commitments drive much of the zeal for a physicalist account of personal identity over time. In looking at our results and how they run so counter to our subjective experience, a physicalist should pause, and not simply write off their experiential evidence. Doing so seems tied to the attempt to, at least theoretically, give third-person physical descriptions of all phenomena, and subjectivity eludes this attempt. Returning to the split-brain operation, we see how a physicalist view of persons treats the first-person perspective as irrelevant. You may wake up from the operation and be looking at a duplicate of yourself, and though others would not be able to distinguish the difference between you and the other body, you could. A physicalist holding a sort of PCC of identity however, would say: "in this situation we should simply claim that the original person is the "same" as both new persons..." However, Moreland and Rae correctly clarify: "they cannot mean…literal identity…What they mean is that sameness is just resemblance…it is arbitrary which one to count as that person [the original, in this case you]" (189). The same move, ignoring the first-person and placing epistemic weight on the third

person, is echoed all over in the philosophy of mind. I continue to maintain that this move is not grounded in facts as experienced, but in ideals driven by metaphysical commitments.[59]

To close, for a dualist: "personal identity is unanalyzable and primitive…The I is ultimate and serves as the unifier of persons…" (180). All our different sensations and thoughts and memories are ours *because* we as individual persons have them. The fact that we have them all is founded on our identity, not constitutive of it. Internal properties of conscious experience like psychological unity and existing as one person through time are the first facts to be taken into account in trying to settle the question of what we are and what grounds our identity. Such experiences, along with others like the intentionality of thoughts and qualia, fit best within a substance dualist perspective and as such they lend yet more reason to adopt a dualist position on human nature over-against a physicalist one which must ignore or seriously distort them (180).

59 Dennett's "Intentional Stance", discussed in an earlier section, is an example of the (in my opinion, unsuccessful) attempt to naturalize the intentionality of thoughts.

## *VI. Physicalism, Dualism, and Bioethics*

With all the objections and insufficiencies of physicalism that we have seen in mind, we can now turn to societal issues to which the question of human nature is potently relevant. There are a plethora of moral decisions to which one's views on human nature can significantly contribute, and since it is individuals who in the end vote on and form public policy, discussing the reasons individuals have for their vote can inform how we make decisions in the public sphere. I'll be drawing from Moreland and Rae's excellent treatment of substance dualism and its implications for ethics in *Body & Soul*, *Human Nature and the Crisis in Ethics*. I'll address abortion, and, more briefly, reproductive technologies, genetic engineering, human cloning (even more briefly), and end of life care.

First, an important distinction made by Moreland and Rae should be explained. In their version of substance dualism, which they call "Thomistic dualism", the soul is more than simply the mind (21). It is the fundamental unifier of the person, that which underlies the development of the organism as a whole, and is metaphysically prior to the development of its parts. Furthermore, there are capacities that persons possess simply in virtue of being persons. These are second-order capacities. The actual ability to develop and use these capacities is a first-order capacities. First-order capacities are not constitutive of personhood, while second-order capacities *essentially* belong to every person just because persons are souls with an intrinsic nature (203). They offer a simple illustration. A second-order capacity one may have is the ability *to learn to speak Arabic*, English, and many other languages. Contrast this with the lower, first-order capacity one may have *to actually speak Arabic* and/or the lack of the first order capacity to speak English (203). Moreland and Rae do not draw this out too far, but one could also talk about the third order capacity to learn languages in general, and the capacity to engage

in cognitive activity as a fourth order capacity, until ultimate capacities are reached (this may be the limit relevant to this example). Finally, one's lower capacities can only be developed in the appropriate environment, but the lack of their development does not entail their absence from a person (i.e. not being able to learn various languages because of being poor and not affording education, or learning language at all because one was, say, isolated from all communication with other persons). This distinction is important in all the morally problematic issues that follow.

### *Human Persons and Abortion*

The current federal legislation on abortion allows (broadly) for abortions past the second trimester and into the third only if justified by a concern for the well-being of the mother. This is contrary to the popular notion that past the second trimester abortion is illegal. In fact, well-being is construed to include psychological, emotional, even financial factors if played right. This was established in the Doe v. Bolton case, a decision made on the same day as Roe v. Wade. Here is the relevant quote:

> We agree that the medical judgment [of the mother's physician, as to whether continuing the pregnancy constitutes a threat to the mother's health] may be exercised in light of all factors–physical, emotional, psychological, familial, and the women's age–relevant to the well-being of the patient. All these facts may relate to health [of the pregnant woman]. (Moreland and Rae 238)

Perhaps less known is that *most* late-term abortions are *not* done for health related reasons, but rather other things that can be construed as "well-being" of the mother, or even of the child (238).

Some have attempted to justify abortion on the grounds that a fetus is merely an extension of the mother's body, and is not a person. The crucial point is that these are attempts to ground personhood. As we saw in the last section, prospects are grim for the continuing identity of persons on physicalism, and these attempts at functionalist accounts of personhood fare no better.

The common strategy of such attempts is to draw a distinction between "human persons" and mere "human beings", where “human beings” are basically the biological organism minus whatever criteria are set-up for personhood. This distinction would have serious implications for positions on personal identity also, though it is usually not thought through enough. If the person is not their body, the biological organism, then somehow mental life is constitutive of personhood, but mental life as an abstract event is not sufficient grounds for identity (Is the process the person? a certain *kind* of mental activity qualifies as a person? or what?). There must be a subject of mental events, and the distinctions made by physicalists we are about to see usually draw heavily on ideas such as intention and self-awareness which seem to be better situated within a substance dualist position.

Prominent thinkers like bioethicist Joseph Fletcher, philosophers Mary Warren, Peter Singer, Helga Kuhse and Bonnie Steinbock all have views of the human person which warrant abortion but which, in the end, are not satisfactory as analyses of personhood. They all rely on the notion of conscious awareness and "having an interest in having interests" (248).

Mary Warren gives the following five specific criteria for personhood:

1. consciousness…and in particular the capacity to feel pain
2. reasoning…
3. self-motivated activity...
4. the capacity to communicate…messages of an indefinite variety of types…

5. the presence of self-concepts and self-awareness, individual and/or racial (245).

Since a fetus cannot, at least to our knowledge, reason, or communicate in the appropriate manner, they are not persons, according to Warren, and abortion is permissible. But, these criteria are not met by the handicapped or infants either. The obvious implication is that killing the handicapped, or infants, could be justified simply in virtue of their not being persons. Warren proceeds to argue against this conclusion because infants are " '*so close*' to being persons that killing them requires additional justification" (qtd. in Moreland, 246), and for the handicapped, there are people who are willing to care for them. But if that's all Warren can offer, then one could simply disagree and say killing something that is not a person is not that morally problematic. The final escape valve for Warren is the argument that an already born baby is no longer a danger to the mother's health. But the extremely small number of cases where this would be applicable in the abortion debate makes it a small help to the logical conclusions of her criteria.

Being somewhat more consistent, Michael Tooley says that infanticide can be justified for up to a week after birth, but that drawing an exact line is not necessary (247). His central criterion is that a person "must have a sense of a continuing self". But how does a purely physical things have a sense of a continuing self? And, when faced with certain psychological disorders, Tooley may find his criteria warranting the killing of persons, like a sever schizophrenia.

Steinbock holds that to be a person one must be able to experience pain or pleasure, and thus have interests (248). For Steinbock also, fetuses cannot have these experiences (until an unspecified point), and thus abortion is permissible. She is inconsistent in her application of her criteria however, since she gives the dead rights (like having their wills carried out), and even says that future persons have rights that current generations have the duty to respect (like

inheriting a healthy planet) (249). The dead and unborn generations cannot currently experience pain or pleasure. Steinbcok disagrees with Tooley's condoning of infanticide by claiming that there is a continuity of identity from the newborn baby to the adult person, but this is pretty much an arbitrary claim since there is no physically distinguishable point during pregnancy that marks the transition from non-person to person.

Further, one can ask pointed questions about all these positions. What do we say about people under anesthesia, or in reversible comas? Is killing them justified? To appeal to the "temporariness" of the condition is to implicitly appeal to the second order capacity of the person to be conscious, and the lack of the first order capacity. If this is done in these cases, the same should be done in the case of a fetus, and the idea that a fetus is not a person can no longer be used to justify abortion. Worse, what do we say about a baby who is born with a lack of higher brain functions, but can be treated to develop them? To appeal to their potential to develop these higher brain functions is to give up the functionalist criteria of consciousness, or an ability to have interests, as grounds for personhood, and instead replace it with some idea of essence, of the capacity *to develop these capacities*, and this is best situated within a substance dualist framework (252).

Because of issues of personal identity which I delineated in the last section, and the problems I have just pointed out, pro-choice advocates have begun to shift the debate from whether or not the fetus is actually a person, to whether it is ever permissible to kill a person. Pro-choice advocate Naomi Wolf puts the issue powerfully:

> Since abortion became legal nearly a quarter-century ago, the fields of embryology and perinatology have been revolutionized, but the prochoice view of the contested fetus has been static…So what will it be: Wanted fetuses are charming, complex, REM-dreaming little beings whose profile on the sonogram looks just like Daddy, but unwanted ones are

> mere 'uterine material'?… abortion should be legal; it is sometimes necessary. Sometimes the mother must be able to decide that the fetus, in its full humanity, must die (qtd. in Moreland 256).

### ***Reproductive Technologies, Cloning, and End-of-Life Care***

The most problematic issue for reproductive technologies is the status of extracorporeal embryos (269). In the most popular methods for enhancing a couple's chances at having children, more embryos are commonly created than can reasonably be brought to term (264). For some processes, such as in vitro fertilization, the excess embryos are frozen in case none of the first four (usually how many are reinserted into the woman's body) are successful, and to avoid having to go through the hormonal treatment needed to harvest the eggs from the woman all over again (264). Also, sometimes there are an excessive amount of embryos implanted, and if too many begin to develop, couples may be referred to another doctor in order to reduce the amount of children they will have (265).

The arguments condoning whatever means are necessary to successfully lead to pregnancy rely on the non-personhood of the extra-corporeal embryos. Usually, the argument refers to the properties that extra-corporeal embryos have *in themselves*, and somehow claiming that they do not have the same higher order capacities that embryos in-utero have (271). More specifically, Singer equates extra-corporeal embryos with the sperm and unfertilized egg before uniting, which cannot develop into human persons unless there is a deliberate human action, and the same, he argues, goes for extra-corporeal embryos (271). Here again however, a substance dualist can reply that egg and sperm become an entirely different entity when joined. Even from a naturalistic perspective, DNA is changed radically, and this is not the case for an embryo that is implanted. An embryo has within itself the potential to develop into an adult person, and is a

person already with such high-order capacities. It will not turn into a different entity, but rather develop to (ideally) display its full potentialities.

Obviously, if one believes that the person's life begins at conception (as a substance dualist will probably believe), then embryos are persons, and one should avoid their deaths if one can. Applied to any couples using reproductive technologies, this belief should engender a more careful use of processes such as in-vitro fertilization (like not fertilizing more eggs than the couple is willing to raise should they all develop) or any other attempts at impregnating. Furthermore, attempts should be made to minimize this in general. One could pursue different technologies (Moreland and Rae list the method of freezing the eggs, instead of the embryos, a method which is still undergoing its experimental phase) (279).

As far as cloning is concerned, the most serious concerns arise with eugenics, but viewing people as embodied persons, whose identity is grounded in the existence of individual souls, doesn't seem to pose any direct objection to cloning. There are already examples of different persons having almost identical DNA, and so long as the technology was not used for unethical purposes there's nothing in the idea itself that causes problems. However, it does seem to be a superfluous technology. One can imagine a plethora of ways it could be used immorally, perhaps because of that potential we should keep our hands off. I personally think that the superfluous nature of the technology is an ethical reason to direct funds elsewhere.

Complications arise again with end-of-life care. Procedures like physician-assisted suicide and euthanasia are the most controversial here, and one's views on *what* a human person *is* are extremely relevant (316). This cannot be determined by empirical investigations since the definitions are not dictated by the facts. For a dualist, patients in a permanent vegetative state (PVS), or in a coma, are still fully persons. For a physicalist, many of the attributes that seem to

constitute personhood have been lost, giving the appearance that one can separate the human organism from the human person. But that separation, forgive the repeating, makes little sense from a physicalist perspective. The distinction being ignored is that one's lack of displaying a first-order capacity does not eliminate the presence of a higher-order capacity.

Perhaps a concern for physicalists is that if we grant full-personhood to such patients, we will not be able to withhold life-support from patients who request it in advance-directives should they one day be in a PVS, or from comatose patients for whom even basic functions like breathing are now being artificially continued, or patients for whom death is imminent and continued treatment will only prolong their suffering. However, this is not the necessarily conclusion of such a premise. Person's wishes should be respected with regards to life-sustaining treatment should they find themselves in a PVS or comatose state. Further, there are situations in which prolonging the life of a patient is indeed to inflict more suffering than is necessary. For such cases there is not a "prolong life at any cost" rule that is wedded to a dualist perspective of persons (337). One thing that would be maintained however, is a resistance to physician assisted suicide, on the grounds that suicide in itself is not good, the real possibility that non-voluntary euthanasia naturally increases with the legality of voluntary euthanasia, and the fact that there are extremely few cases where pain relief cannot be adequate. In this vein, Moreland and Rae acknowledge that there need to be looser reigns on doctors' abilities to administer pain-relieving drugs like morphine in higher doses so that patients who are terminally ill can die without unnecessary suffering (341).

Regardless of one's position on any of these issues, it is obvious that *the* central issue is what constitutes personhood, and I am continuing to argue that substance dualism does the best job of accounting for personhood.

# VII. Some Objections to Dualism

We have seen that there are many objections to physicalism. Thinking about the gravity of some of them spurs the question of why one would not take up dualism as the most probable truth about human nature. Well, the philosophically relevant reason is that there are also objections to dualism which seem, to some, to have weighty significance. I'll address some of these objections now, with as much clarity as is possible in a limited space. Most of the philosophers from whom I derive these answers have written more extensively though, and I encourage an exploration of the rigorous defenses.

### *What's the Soul Made Up Of?*

The first objection is expressed well by Michael Levin who writes:

> The trouble, I suggest, is this: we can say what sort of stuff a material thing is an individual piece of, while no one has any idea of the sort of stuff a self is an individual piece of…While there are descriptions that can identify a self, we cannot refer to it as a P of S [it is not a part of something else] (quoted in Plantinga, *Materialism* 121).

Plantinga deals with this objection fairly effectively, I think. The essence of the reply is that, if this objection is supposed to seriously move someone from dualism to physicalism, or in that direction, it would also cast doubt on the existence of many other things we believe exist (124). Think of a property, a proposition, or a set, or a number, or an electron, or the quantum field that the electron is a part of. What is a property made up of? What is a proposition made up of? What is a set made up of (its members obviously, but the set itself does not seem to be a physical reality)? The questions become more puzzling. What are numbers made up of? What are top quarks made up of? That last question affects the physicalist also. As contemporary physics digs further into the bedrock of material reality, the question of what things are made up

of never seems to end, and the ultimate answer starts to transform into theoretical physics and entities whose composition we have no idea about (strings?). Does this count against the reality of matter itself? Does everything need to be made up of some other stuff?

The answer to that last question is, in my opinion emphatically no. To maintain that idea seems incoherent anyway and threatens an infinite regress of things composed of other yet other things. What is more, a dualist believes that a person is a simple entity, not composed of anything, and that has been the line taken by almost all dualists. So it is in the end unclear to me how this objection is supposed to move a dualist. If one takes stock of all the problems of physicalism, feels like they should start to investigate dualism, and runs up against this objection or one similar to it, there should be no problem.

***Dualism Is Anti-Scientific***

Along the same lines, materialists often think of the existence of a non-material soul as a scientific postulate, one that is "designed to explain the phenomena, something that gets whatever warrant it enjoys by virtue of the excellence of the explanation it provides" (Plantinga, *Materialism* 124). I myself think that there is a viable way to defend dualism as a scientific hypothesis, and there is research underway that I believe will enable a dualist to contribute to the discussion in this way and discuss it in the next few sections. On the other hand, perhaps to the distaste of a strict empiricist, one can believe that human persons are souls because one believes that a purely physical thing cannot think, or have qualia, but human persons do think; or because purely physical things do not maintain their identity through time, whereas human persons do. The empirical data leave the question more than just open; they cast doubt on the physicalist's stance because of *its* explanatory failures. To take a different line, one could say that dualism does, at least, offer an alternative explanation of phenomena that physicalism gives implausible

accounts of. For example (again), how it is that humans think, or are conscious, or maintain identity through time?

In particular, the idea that cognitive science is antithetical to dualism is loudly proclaimed. I personally deny this and affirm, as most dualists would, the intimate link between body and soul. Some think that the fact that damage to certain parts of the brain results in mental malfunction is a problem for the dualist, but I've never been able to makes much sense of the suggestion. In our embodied lives, interaction with the material world is mediated, even if only with a thin veil, and if the physiological structures by which we interact with the world are damaged, it only makes sense that the activity reliant upon them would be adversely affected. That the same occurs with mental diseases is still no problem for dualism. The brain is used to organize inputs and final outputs of our mental lives, and when the neurological order that these are founded on is in disarray, one's cognitive life will undoubtedly be the same. One can accept the correlations between loss of memory, or vision and the corresponding damage to parts of the brain. Some dualists are scientists who have significantly contributed to relevant fields, take Wilder Penfield's in his The Mystery of the Mind and John Eccles Facing Realty: Philosophical Adventures by a Brain Scientist (Plantings, *Materialism* 123).

### *Mind Kiss Matter?*
### *The Energy Conservation and Interaction Objections*

Similarly, many thinkers feel that if an immaterial mind were to interact with the physical world, say with the brain or body, this would violate some fundamental law of physics, and wreak havoc in the physical universe (125). This is commonly dubbed the "interaction problem". *How* an immaterial mind could interact with the physical world, seems to present such difficulties that some physicalists tout it as a deathblow to dualism. Dennett expresses the problem as follows:

> A fundamental principle of physics is that any change in the trajectory of any physical entity is an acceleration requiring the expenditure of energy, and where is this energy to come from?…This confrontation of quite standard physics and dualism has been endlessly discussed since Descartes' own day, and is widely regarded as the inescapable and fatal flaw of dualism. (qtd. in Plantinga, *Materialism* 125)

Offhand, a dualist might say that the law of conservation of energy applies to closed systems, and if souls exist, then ex-hypothesi the body is not a closed system (*Materialism* 126). As concerns the mind's causal power upon the physical, the objection is seldom supported by serious argument, but rather stated with rhetorical flourishes, based on the assumption that this interaction would be utterly inexplicable. However, on this point there are no empirical data that count against the possibility of an immaterial mind "causing" events in the physical world, and the arguments already discussed thus far should put the burden on the physicalist to explain how a physical thing could have a mental life like ours that is causally efficacious. Somewhat embarrassingly to this objection however, there are responses within our knowledge of the physical world itself that completely strip this objection of its perceived force.

The objection in general is based on two assumptions. They are

1. that the principle of energy conservation applies to all known purely physical interactions and
2. that all causal interactions (or law-like connections) between events must involve an exchange of energy (Collins 125).

I will be quoting heavily from Robin Collins' essay “The Energy of the Soul” to illuminate the Dualist response. We start by trying to understand how energy is measured in physics. Collins begins:

> In classical mechanics, the total energy of an object is equal to the sum of the internal energy of a body and its kinetic energy, $1/2mv^2$. The latter quantity, however, will depend on the frame of reference from which the velocity of the object is measured. In special relativity, the frames of reference of interest are those moving at some uniform (non-accelerating) speed relative to each other. These are called *inertial* frames of reference. (126)

An example of an inertial frame of reference that Collins gives involves a train moving at the same speed with the ground, or, put more interestingly, a train moving at the same speed as a ball next to it (say the ball is moving on top of a car driving alongside the train at the same speed). Calculated from the train, the ball has no kinetic energy. But, calculated by people standing on the ground (while they themselves and the ground are both moving at the same speed) the ball and train will both have values for v in the kinetic energy formula (126). The end result of this basic feature of measurement is that "...unless we establish a preferred frame of reference, we cannot speak of the energy of the ball as being an intrinsic, non-relational property of it" (Collins 126).

Collins draws out what this means for general and special relativity, the current framework in which we approach measurements of energy. "A central idea behind both the special theory of relativity and the general theory of relativity is that the laws of physics should be formulated in terms of quantities that are independent of one's frame of reference" (126). Since energy is not frame-independent, both special and general relativity posit replacement quantities. For special relativity, "the frame independent quantity that substitutes for energy is a four component entity called the energy-momentum vector" (126). For general relativity, this entity is "a sixteen-component tensor called the stress-energy tensor" (127). The energy-momentum vector in special relativity and the stress-energy tensor in general relativity "can be

considered an intrinsic property of an object or system of particles and fields, but its energy cannot"(127). These quantities would serve us in giving some constant value for energy by which to measure energy and thus energy conservation, but complexities arise.

Given the idea of frames of reference, and knowing that calculating the energy[60] of a system itself is relative to a frame of reference, the energy conservation principle in physics, that Collins gives, is the "boundary version". This version states: "the rate of change of total energy in a finite region of space is equal to the total rate of energy flowing through the spatial boundary" (127). So, one would subtract however much energy is being lost by the system from however much energy is being put into it. Calculating this would incorporate the stress-energy tensor, and this would result in a consistent way to formulate the energy conservation principle.

The problem for the boundary version of energy conservation is that it does not hold for any region of our universe. This is because of gravitational energy; Collins explains, "typically one can neither define the total gravitational energy in a region of space nor the rate at which gravitational energy flows in or out of the region" (128). A reason for why we cannot do this is that we cannot calculate a stress-energy tensor for gravitational waves, and thus cannot formulate a "frame invariant quantity", for gravitational energy or momentum. But, gravitational waves are constantly affecting all the matter in the universe. We can actually record the increase in non-gravitational energy within a system without being able to identify any "physically definable energy flowing across the boundary of the region" (129). In other words, there's no precise method of measuring the energy coming in and out of a given local region of space without ignoring gravitational energy. This suggests that since there is no well-defined local conception of energy conservation, there is no global conception either, since gravity affects all matter and

60 I'll copy Collins in just using "energy" now instead of "stress-energy".

there's no precise way to define the energy for gravitational waves.[61] Pushing this further, Collins quotes philosopher of science Carl Hoefer's suggestion that because we have no trace of its source, or measurement of the total amount of gravitational energy pervading our universe, "energy gain in a gravity wave detector could be thought of as genuine gain, without having to say that the energy existed somewhere beforehand" (qtd. in Collins 129).

Moving on to the problem of causality, there are established cases in quantum mechanics where physical particles interact without any exchange of energy. Collins uses "particle spin" to illustrate this point. In quantum mechanics all particles have a spin with a value of +1, -1, +1/2, -1/2, or 0, and this value never changes except as being either positive or negative. If we split a nitrogen molecule (composed of two nitrogen atoms), which has a spin of zero, in a spaceship midway between Earth and Mars, send each atom to a separate planet, and then have someone on each planet measure their spins in a pre-determined direction, each person will either measure a spin of +1/2 or -1/2, and what their measurements are will be anti-correlated (131).

One naturally thinks that each particle's spin was determined when they were split, and then the instruments measured the results. If this explanation were correct, it would be explaining via only "local causation". When the atoms were split, there was no significant spatial distance between them and the mechanism that gave them that spin. Later, when their spins were measured on different planets, there was no significant spatial distance between them and the apparatus that registered their spin (so their spin state caused the reading). I cannot possibly illuminate the full-story to rebut this, but it is has been well established via several different

61 Collins also points out the energy loss that most of the photons in the universe undergo, which we detect in cosmic redshift and rely on to calculate the rate of expansion of the universe and how far stars are. The energy lost by photons however, seemingly goes nowhere. Collins backs up the math of his argument extensively and will point the reader to the relevant physicists who have expressed the problem

experiments all under the umbrella of "Bell's Theorem" that *local causation is not sufficient to explain the anti-correlation*[62].

But, all energy exchange involves local causation, so the correlations cannot be explained via energy exchange. The two responses to this fact in quantum mechanics are categorized as "causal realist" and "causal anti-realist" interpretations. The causal realist response is that there is some "instantaneous causal connection between the two particles or in some non-local and thus instantaneously acting common cause" (132). The causal anti-realist response is that these correlations cannot be explicated further (132). If you take the realist response, you're accepting that there are causal connections that don't require energy-exchange. If you adopt the anti-realist response, then requiring that forms of dualism that posit law-like correlations between the mind and the brain to explain their correlations further is to ask of immaterial-material interactions something that not even physical-physical interactions provide (132).

These two results, the inability of physics to delineate a consistent energy-conservation principle or a theory of physical causation which requires energy exchange, completely strips the energy-conservation principle objection against dualism and casts doubt on the force of objections directed at the implausibility of causation without energy exchange.

62 I refer the readers to the Stanford Encyclopedia of Philosophy's article on "Bell's Theorem" for their own personal confirmation.

# *VIII. A Dualist Interpretation of Quantum Mechanics*

Philosopher of Physics Hans Halvorson, in his essay *The Measure of All Things: Quantum Mechanics and the Soul*, delineates how a dualist perspective on human nature can help explain some paradoxes that have arisen over the past several decades concerning the behavior of the universe's elementary physical particles. Specifically, Halvorson outlines how Dualism can contribute to our understanding of the "measurement problem" and the collapse of the "wave-function". I'll first explain the concepts needed to understand the problem.

**Quantum Superpositions**

First, there's the idea of superposition. Coming into the 20th century, physicists devised an experiment to determine whether the newly discovered negative charge on the outer skirts of an electron is carried by waves or by particles. The test was to send negative charge to an optical detector screen from some source, but to place a barrier with two doors in-between the source and detector, and, like all scientific experiments, to change the independent variables and see what happened (143). If negative charge is carried by particles, then there should be discrete lumps on the detector behind each door signifying where the particles hit (see diagram 1 on next page). If negative charge is carried by waves, then there will be a diffuse mark on the whole detector, plus a specific pattern indicating where the waves interfered with each other if both doors are open (diagram 2 on next page).

**Diagram 1: Two-slit experiment (results for particles)**

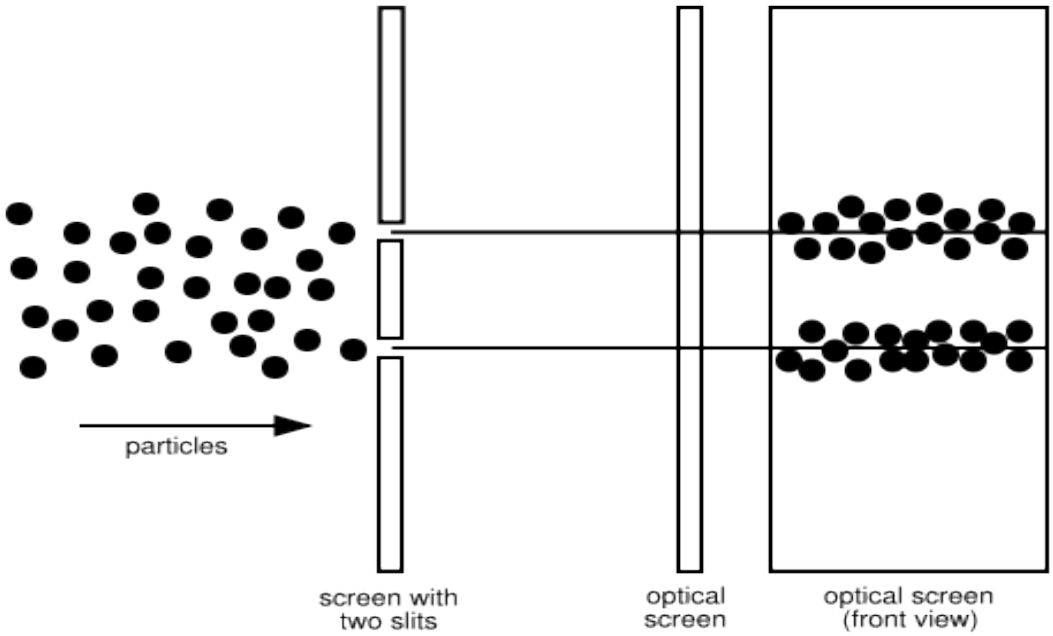
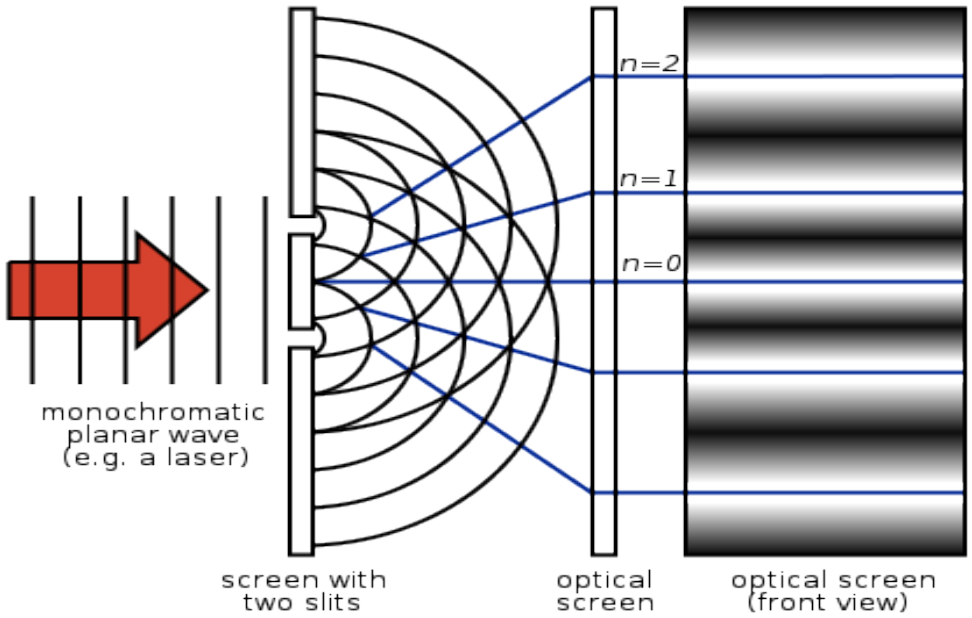



**Diagram 2: Two-slit experiment (results for waves)**




(diagrams obtained from wikibooks "Materials in Electronics/Wave-Particle Duality/The Two-Slit Experiment")

Now, say we keep the negative charge source on *and* keep both doors open. If negative charge is carried by particles we should expect for there to be two lumps, one behind each door. Instead, we see a diffuse pattern on the optical screen, like in Diagram 2. However, there had already been experiments giving *strong* empirical evidence that negative charge was carried by particles, electrons, but this experiment contradicted that. Only waves can create that interference pattern, it so it seemed as though negative charge was carried by waves. Even when one shoots only the minutest amount of negative charge at a time (one electron at a time), the diffraction pattern still appears, and this seems to be undeniable evidence that negative charge is carried by waves. However, skeptical of this result and wanting more confirmation, another experiment was devised which placed a detector over each door and see if both detectors went off, as should happen:

> the result is surprising: in any particular run of the experiment when the detectors are turned on exactly one detector goes off (confirming that electrons are localized particles). But when these detectors are turned on [and both doors are open], the interference pattern on the optical screen disappears, and instead we get the two-lump pattern on the optical screen. It is as if the electron behaves like a particle in the presence of the detectors, but like a wave when there are no detectors (Halvorson 143)

So, the physicists who made these discoveries "invented a new concept called the 'quantum superposition of two states' and they claimed that when both doors are open…the electron is in a quantum superposition of passing through the left and right doors" (Halvorson 144). The notation for this state is: "|left>+|right>". This means that until we detect it, the electron *does not have a position, period*. Here we make an important note, one *cannot* think that this superposition state is the result of our ignorance about which door the electron will go through, because when there are no detectors the electrons produce a diffuse pattern on the

optical screen which would not make sense if they did in fact acted like particles! However, when we do make try to find the electron with detectors, the superposition state |left>+|right> does work like a probability. Half the time we'll detect it in the left door, the other half of the time we'll detect it in the right door (144). So, electrons act like waves with no localized position in space. But when measured, they act like particles. This is what physicists call the "collapse of the wave function" (144).

Collins explains that *all* quantum states are superpositions. The state |left> in our experiment is itself a superposition of two other states:

> |moving>…
> in which the electron is moving, and the state
> |stationary>…
> in which the electron is stationary (145)

The state |left>, remember, meant that we were seeing the electron pass through the left door via a detector. What the superposition above means is that, should we detect the electron going through the left door, we could find its position or its velocity, but not both. An electron is not stationary (not in a position) when moving, nor does it have a velocity (its not moving) when in a position, and neither until we make a measurement[63]. An electron *always* is missing some indeterminate values that normal physical objects should have, it's always in a superposition. This leads to the next concept, "entanglement".

**The Snow Leopard's not Dead**
**Entanglement and Linear Dynamics**

Suppose two electrons, Allie (A) and Bryna (B), are going to be in two separate two-slit experiments (146). Quantum mechanics can describe the combination of both states as:

$(|left\rangle_A + |right\rangle_A)(|left\rangle_B + |right\rangle_B)$

63 I leave it to the reader to do further research in how this is so, there are a plethora of non-technical books on the basics of Quantum Mechanics.

or, validly applying the distributive function:

$|left\rangle_A|left\rangle_B + |left\rangle_A|right\rangle_B + |right\rangle_A|left\rangle_B + |right\rangle_A|right\rangle_B$

This last state describes all the possible outcomes for the measurements taken of the two electrons. But remember, it is a *superposition*. Therefore, neither of the electrons is *actually* in these states until the wave function collapses (147).

It has been established through a series of experiments dubbed "Bell's inequality" that not only can two states of an individual electron be in a superposition, but also two states that are each made up of two electrons. So, let's say that the two electrons are in the superposition state: $|left\rangle_A|left\rangle_B + |right\rangle_A|right\rangle_B$. Halvorson short hands this state $|E\rangle$, and I'll follow suit. $|E\rangle$ says that "if we perform a joint position-position measurement, we will always get the same result for both electrons" the two electrons always go through the same door. (147).

Entanglement is the realization that in state $|E\rangle$ , neither A or B have *any determinate properties at all* (148). This is because A cannot be in state $|left\rangle$, since $|E\rangle$ says it could go through the left door. A is not in the state $|left\rangle + |right\rangle$, because this would mean that A may go through a door that B does not go through, but that's not possible in state $|E\rangle$. So, A is not in any state, Halvorson says that A's "quantities – position, velocity, etc. – lack determinate values" (148).

Entanglement is a problem because *everything* is made up of elementary particles that obey the laws of quantum mechanics, and there have been no experiments indicating a "cut-off point" (148). Say you take two particles that are parts of the heart of a snow leopard together with 5,000 more particles in their vicinity, and describe the state of that "composite system"

(148)[64]. All the elementary particles will have been in superpositions just like A&B in our earlier example, and so when you describe the state of the resulting system, you will have to list all the possible combination of states it may be in. In other words, the system is in a superposition of states, and one could theoretically continue this process all the way up until one describes the state of the entire snow leopard, resulting in a ridiculous amount of possible states, even the superposition of |dead>+|alive>, in which our feline "has no state at all, is neither alive nor dead, is not awake or asleep, etc." (148). One can now see a hint of what the problem will be. Our experience seems to deny that macro-objects are in superpositions of states, at any given moment the snow leopard seems to be *either* alive *or* dead, *either* asleep *or* awake, but not both.

The last concept needed to understand what will be the most troubling consequence of quantum mechanics, is linear dynamics. Classical physics used to postulate deterministic laws for how deterministic systems would change over time; quantum mechanics postulates a similar law (the Schrodinger equation) according to which whatever the state a quantum system is *now* will determine its future quantum state. The problem is that superpositions are maintained through change. Halvorston states:

> If the state |S> were to evolve into the state |S'>, and if the state |T> were to evolve into the state |T'>, then the superposition state |S>+|T> would evolve into |S'+|T' (149).

The "measurement problem" is a troubling description of the physical world if linear dynamics, the existence of quantum superpositions, and entanglement are true.

64 This is an over-simplified version of the notorious "Schrodinger Cat"

**Not So Super:**
**The Measurement Problem**

Halvorson paints the following scenario and quotes philosopher David Albert to describe what will be the concluding dilemma of our excursion into quantum mechanics. Imagine we set up the two-slit experiment we described earlier and attached a computer monitor to the detectors. The state of the electron before the experiment, remember, is the superposition: $|left\rangle_E+|right\rangle_E$. We could label the state of the computer monitor before the electron is shot as: $|ready_M\rangle$. So, we could describe the combined state of the computer and the electron as the superposition: $|ready_M\rangle(|left\rangle_E+|right\rangle_E)$. This is the same as: $|ready_M\rangle|left\rangle_E+|ready_M|right\rangle_E$. If the left detector goes off, the computer screen will display "left", and it will display "right" if the right detector goes off. Now, let's say the electron is shot. According to linear dynamics, superpositions are preserved, so the superposition of the monitor-electron system will evolve into another superposition: $|left_M\rangle|left\rangle_E+|right_M|right\rangle_E$. But, recalling our discussion of entanglement, this means that the monitor will not display left or right, because it could display the other depending on what the electron does, and it cannot be in a superposition, because the electron is what determines what it's supposed to read, but the electron is still in a superposition! So, the monitor will not read anything at all because any state it could be in results in a contradiction with the predicted result of quantum mechanics!

Add a human observer to this weird mix. Desmond is looking at the monitor before the electron is shot. The state of the composite system of Desmond, the monitor and the electron is a superposition, namely:

$|ready_D|ready_M\rangle|left\rangle_E+|ready_D|ready_M|right\rangle_E$

Now, if the monitor comes to display "right" Desmond will believe that it says "right", he'll be in the state: $|right_D>$ If the monitor comes to display "left" Desmond will believe it says left, he'll be in state: $|left_D$ However, remember what happened with the computer. Assuming that Desmond is, like the monitor, a reliable detector, his state is dependent upon the final state of the monitor, but the monitor will evolve according to linear dynamics! So, the final state of the Desmond-monitor-electron system will be:

$|left_D|left_M>|left>_E+|right_D|right_M|right>_E$

This means that Desmond will not believe anything about what the monitor says.

I repeat, quantum mechanics, what Halvorson dubs "the best physical theory in history", predicts that when Desmond, or any human being, makes an observation on the monitor, or any observation, they will be in a quantum superposition in which they will observe neither one option in the scenario nor another. Our experience demonstrably falsifies this prediction. When we make an observation, we observe one thing, not its contradictory, and we certainly do observe something. We have definite experiences of both sensory observation and belief formation that are not in superpositions (I do not believe both that the traffic light has turned green *and* that it has not). This is recognized as a *serious* problem. *So much* of a problem that great energy and ingenuity has gone into saving the theory of quantum mechanics by rejecting one of its assumptions and/or developing an alternate theory of the universe as a whole.

We will not delve into all these different methods, but I refer the reader to two books Halvorson suggests for varying interpretations. One is D.Z. Albert's *Quantum Mechanics and Experience*, the other is Jeffrey Barrett's *The Quantum Mechanics of Minds and Worlds*. I will briefly state how dualism, the thesis that humans are immaterial minds, souls, and *not* purely physical entities can resolve the problem.

**Soul Scientists:**
**The Dualist Interpretation**

In the preceding chapters, we noted how a physicalist view on human nature leads to varying difficulties, some tolerable, some, I argue, insurmountable. Let's say that we consider dualism a viable option now. Given that possibility, what would one say to the measurement problem? Well, return to Desmond and the state composed of himself, the monitor, and the electron (DME hereafter). According to quantum mechanics, the components of this system will all evolve into a superposition in which the monitor and Desmond are entangled with the electron and each other. Thus, the monitor does not display "right" or "left", and Desmond does not observe anything. The dualist throws a wrench in the problem because, according to a dualist, mental states are not made up of the elementary physical particles that govern the behavior of quantum dynamics, and thus, there's no prima facie reason to think that mental states are superposable[65]. We saw at the beginning of this discussion that the reason physicists invented the idea of a quantum superposition is that it had explanatory power; it explained how an electron could produce a diffraction pattern in the two-slit experiment. There is no observational data that the "superpositions of mental states" would help explain. If anything, it contradicts observational data. However, if mental states cannot enter into superpositions, they cannot become entangled, and so, though the state of (DME) before the electron goes through a slit in the experiment is:

$|ready_D|ready_M>|left>_E+|ready_D|ready_M|right>_E$

and though this is supposed to result in an entangled state for both the monitor and Desmond, this postulate can be said to not apply in situations where the components of the system are not purely physical. Again, what phenomenon would the superposition of mental states explain?

65 Halvorson cites Chalmer's suggestions in *The Conscious Mind* as intimating that this is the solution to the measurement problem that people should be taking into consideration.

Hans Halvorson cites physicist Henry Stapp's recent book *Mindful Universe: Quantum Mechanics and the Participating Observer* and his intriguing essay *Quantum Theory and the Role of Mind in Nature,* as more detailed scientific expositions of how individual conscious subjects play important explanatory roles in our contemporary understanding of the universe, and why this needs to be recognized not only in physics, but in psychology and neuroscience as well. One can look into Stapp's other works and lectures to glean the arguments and well-supported evidence for his position. Additionally dualism can contribute to the already existing "Ghirardi-Rimini-Weber collapse theory" (GRW) according to which there are *extremely rare* spontaneous collapses of the wave function. But, there are so many particles in even what we would call a "small" object that this happens frequently enough. Since two particles that are entangled will both collapse if one does, all the particles that compose macro-objects collapse, and the object will obtain definite properties, just as we observe them to have. Dualism can contribution to our understanding of these spontaneous collapses of the wave function. Halvorson says that in GRW, "the collapse dynamics seems to be ad hoc, and put in by hand to solve the measurement problem" (162). For a dualist, these theorized collapses would not be random and spontaneous, but directed by conscious decisions of immaterial minds under indeterministic laws governing brain-body interaction (Stapp provides frameworks for how to develop this). The dualist perspective, undergirded by independent evidence, can underwrite the development of this theory and give it grounds.

One may reply skeptically at first. Is not the postulation of the soul, or an immaterial mind simply ad-hoc in this situation? Well, no. As I've argued throughout this thesis, dualism makes the most sense of our experience *prima facie*, and there are many reasons not to be a physicalist. Now we've come to a physical theory that can escape the charge of contradiction and

incoherence if interpreted with a dualist, rather than a physicalist, metaphysics. From what I've seen in my research, this trend is on the rise, rather than fading away.

# *X. Conclusion*

I want to take a step back and mention my original rationale for writing this thesis. The desire was undergirded by many shifts in my personality, beliefs, and interests that have occurred in the three years I've been an undergraduate. The two most important are my conversion to Christianity before my sophomore year, and a newfound, very intense desire to discover reality (what is true about the universe and our place in it?). One could oppose this desire to discover reality to a mindset that says: "Well, maybe what I believe is false, but I'm comfortable believing it." It is easy to see how a false view of reality can put one in danger; imagine jumping from a helicopter while trusting a parachute that, in reality, is defunct. I admit, this danger is not so easily seen when it comes to some of our other beliefs, but our conception of what we are underlies virtually everything else we do. To use a phrase I'm not so fond of, it's a belief that "cashes out" so much that it's hard to trace the paper trail back to it.

I have presented arguments directed towards the physicalist enterprise as a whole. The second section was probably least persuasive to anyone who already was a physicalist, but I extended Plantinga's example to include a possible wrinkle in the data collection that would require a physicalist to abandon the naive principle "just following the evidence". The third and fourth sections were connected in their question for the physicalist: how does physicalism make sense of our mental lives? My fifth section was all too brief for the topic, but I refer the reader to my source, and I think that Plantinga's case, though it runs counter to some modern assumptions about evolutionary biology, is actually very potent in that it attacks the *source* of *all* our judgments. I also pointed out why the personal identity of a human being is seated in the mind, which possesses what Moreland and Rae call an "individuating principle". The implications for this in ethics can be revolutionary. Of course, we would need a serious paradigm switch, but

nothing's to say that such a switch is impossible. My last two sections are meant to intrigue. One has to do the research and reach their own conclusions. But its aim is to point out that both scientists and philosophers are taking dualism as a serious possibility. This indicates there is still openness in the debate, and that is encouraging.

Over the past nine months, the research noted here has put me in a position where the most rational thing for me to do is to think that human consciousness and cognition are not solely physical enterprises. Whatever aspect of this thesis stirred the most curiosity is what I encourage others to look at more carefully. But, that much is necessary, some *effort t*o try and make sense of the views at hand. At the least, I hope these ruminations are exercises that sharpen my and future reader's abilities to get-at what the truth is, not only for the question of human nature, but also for our understanding of the universe as a whole.

# Works Cited

Battaglia-Mayer, Alexandra and Roberto Caminiti. "Optic ataxia as a result of the breakdown of the global tuning fields of parietal neurones." Brain 125.2 (2002): 225-237.

Bermudez, Jose. "Arguing for Eliminativism." Paul Churchland. Ed. Brian L. Keeley. New York: Cambridge University Press, 2006. 32-65.

Boghossian, Paul A. "The Status of Content." The Philosophical Review 99.2 (1990): 157-184.

Brown, Curtis. "Narrow Mental Content." 2008. Stanford Encyclopedia of Philosophy (Winter 2008 edition). 2011 <http://plato.stanford.edu/archives/win2008/entries/content-narrow/>.

Churchland, Paul. "Eliminative Materialism and the Propositional Attitudes." The Journal of Philosophy 78 (1981): 67-90.

Collins, Robin. "The Energy of the Soul." The Soul Hypothesis Investigationg into the Existence of the Soul. Ed. Mark Baker and Steward Goetz. New York: The Continuum International Publishing Group Inc. , 2011. 123-137.

Crisp, Thomas M. "Gettier and Plantinga's Revised Account of Warrant." Analysis 60.1 (2000): 42-50.

Dennett, Daniel. The Intentional Stance. Cambridge: MIT Press, 1996.

Dretske, Fred. Naturalizing The Mind. Cambridge: MIT Press, 1997.

Halvorson, Hans. "The Measure of All Things: Quantum Mechanics and the Soul." The Soul Hypothesis: Investigations into the Existence of the Soul. New York: The Continuum Publishing Group, 2011. 138-167.

Horgan, Terence and James Woodward. "Folk Psychology is Here to Stay." The Philosophical Review 94.2 (1985): 197-226.

Inwagen, Peter Van. Metaphysics. Boulder: Westview Press, 2009.

—. "Plantinga's Replacement Argument." 2007. Department of Philosophy//University of Notre Dame. 2010 <http://philosophy.nd.edu/people/all/profiles/van-inwagen-peter/documents/PlantRplArg4.doc>.

Madell, Geoffrey. Mind and Materialism. Edinburgh: Edinburgh University Press, 1988.

—. "The Road to Substance Dualism." Royal Institute of Philosophy Supplements 67 (2010): 45-60.

Moreland, J.P. and Scott Rae. Body & Soul: Human Nature & the Crisis in Ethics. Madison: InterVarsity Press, 2000.

Plantinga, Alvin. "Against Materialism." January 2006. University of Notre Dame Department of Philosophy. <http://philosophy.nd.edu/people/all/profiles/plantinga-alvin/documents/AGAINSTMATERIALISM.pdf>.

—. "Content and Natural Selection." University of Notre Dame Department of Philosophy. <http://philosophy.nd.edu/people/all/profiles/plantinga-alvin/documents/CONTENTANDNATURALSELECTION.pdf>.

Plantinga, Alvin. "Introduction: The Evolutionary Argument against Naturalism." Naturalism Defeated? Essays on Plantinga's Evolutionary Argument against Naturalism. Ed. James Beilby. Ithaca: Cornell University Press, 2002. 1-14.

Plantinga, Alvin. "Materialism and Christian Belief." Persons: Human and Divine. Ed. Peter Van Inwagen and Dean Zimmerman. New York: Oxford University Press, 2007. 99-141.

Plantinga, Alvin. "Reply to Beilby's Cohorts." Naturalism Defeated? Essays on Plantinga's Evolutionary Argument against Naturalism. Ithaca: Cornell University Press, 2002. 204-276.

Putnam, Hilary. "Brains in a vat." Knowledge: Readings in Contemporary Epistemology. Ed. Sven Bernecker and Fred Dretske. New York: Oxford University Press, 2000. 385-399.

Ravenscroft, Ian. "Folk Psychology as a Theory." 2010. The Stanford Encyclopedia of Philosophy (Fall 2010 Edition). Ed. Edward N. Zalta. <<http://plato.stanford.edu/archives/fall2010/entries/folkpsych-theory/>>.

Taliaferro, Charles. Consciousness and the Mind of God. New York: Cambridge University Press, 1994.