



1996

# Two Problems in the Philosophy of Mind

C.J. Summers '96

*Illinois Wesleyan University*

---

## Recommended Citation

Summers '96, C.J., "Two Problems in the Philosophy of Mind" (1996). *Honors Projects*. Paper 2.  
[http://digitalcommons.iwu.edu/phil\\_honproj/2](http://digitalcommons.iwu.edu/phil_honproj/2)

This Article is brought to you for free and open access by The Ames Library, the Andrew W. Mellon Center for Curricular and Faculty Development, the Office of the Provost and the Office of the President. It has been accepted for inclusion in Digital Commons @ IWU by the faculty at Illinois Wesleyan University. For more information, please contact [digitalcommons@iwu.edu](mailto:digitalcommons@iwu.edu).

©Copyright is owned by the author of this document.

# **Two Problems in the Philosophy of Mind**

**C.J. Summers**

5/20/96

Honors Research Project  
Advisor: Dr. Lenny Clapp  
Illinois Wesleyan University

Two Problems in the Philosophy of Mind

By

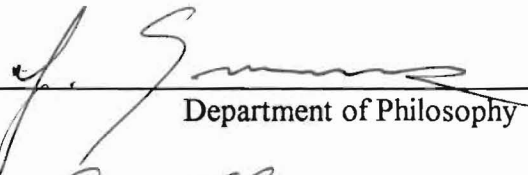
Carl John Summers

May 20, 1996

Research Honors Project

Illinois Wesleyan University

Signature of Author



Department of Philosophy

Certified By



Project Advisor

Certified By \_\_\_\_\_

Certified By \_\_\_\_\_

Certified By \_\_\_\_\_

## Table of Contents

<b>Introduction</b>	<b>1</b>
<b>Part I: A Criticism of Jerry Fodor's Theory of Mental Content</b>	<b>3</b>
1. Why Mental Processes are Physical	4
2. Why Mental States are Intentional	5
3. The Big Problem	7
4. Fodor to the Rescue with the Big Problem	8
5. The Little Problem	9
6. Fodor to the Rescue with the Little Problem	11
7. Block's Obstruction	12
8. Fodor Overlooks Perception, a Revision to AD	14
9. Small Green Cows Pose No Threat to Fodor	20
10. How Fodor Fails	21
11. The implications of Fodor Failing	26
<b>Part II: A Criticism of Daniel Dennett's Theory of Consciousness</b>	<b>29</b>
1. Dualism Discarded	30
2. The Persistence of Our Confused Intuitions	30
3. What Consciousness Really Is, the Multiple Drafts Model	32
4. Evidence Against a Central "I," the Color Phi Phenomenon	33
5. Two problems With Dennett's Theory	37
6. Where this leaves Dennett	39
<b>Conclusion</b>	<b>41</b>
<b>References</b>	<b>43</b>

Today, the opinion prevails that everything can be "rationally explained", as they say, by alert reason. Nothing is obscure - and if it is, then we need only cast a ray of scientific light on it and it will cease to be so.

This, of course, is only a grand self-delusion of the modern spirit. . . . At most it will drive the complex world of hidden things to find surrogate, counterfeit, and increasingly confusing manifestations. . . (Vàclav Havel p.159-160)"

## **Introduction**

The philosophy of mind has historically been concerned to a large part with two central phenomenon of human experience. The first is the intentionality of our mental states, the fact that they seem to be meaningful. The second is the fact that humans are conscious beings. Not only do we receive and process information, we seem to be aware of the experiences which constitute our input and are cognizant and in control of many of the processes which are performed upon this information. It is of obvious interest how we manage to have intentional mental states and be conscious beings. Traditional discussion of these phenomena has usually relied upon the postulation of a mind which implies a dualist ontology. The presumption of a mind, which is composed of non-physical, mental, substance makes these problems relatively easy to solve. Mental substance which is by nature unobservable can serve just about any purpose which we can imagine for it, including possessing meaning and constituting consciousness.

The advent of what Vaclav Havel calls a "scientific world-view" has brought disrepute to explanations which appeal to a dualist ontology. The successes of science as an explanatory tool, unraveling mystery after mystery, has lead to the assumption that there is in fact no mysteriousness, that everything is common and explainable in common scientific terminology. The philosophical

formulation of these attitudes is the theory of physicalism. Physicalism claims that everything which exists is composed of physical matter, and specifically denies the existence of anything like the dualist's mental substance. This of course resurrects the problems of intentional content and consciousness. It was relatively easy to explain them by appeal to mental substance but it becomes much more difficult when the tools of explanation are limited to those allowed by physicalism.

At this point there are two obvious positions which one may take regarding the phenomena of intentional content and consciousness. It is possible to deny that they exist at least in the sense in which our intuitions lead us to believe. The other option is to show how these phenomenon are in reality purely physical, how they are explainable in terms allowed by our scientifically based physicalism. This strategy is known as naturalization and has been adopted by Jerry Fodor in the case of intentionality and Daniel Dennett in the case of consciousness. They each claim to have provided physical explanations for these seemingly mysterious facts about human existence. The following paper presents a criticism of their respective efforts at naturalizing intentional content and consciousness.

## **Part I: A Criticism of Jerry Fodor's Theory of Mental Content**

Jerry Fodor is one of the premiere philosophers of mind alive today. As any serious philosopher of mind must, he has a theory regarding intentional human mental states and their conflict with physicalism. Fodor is a realist about human intentional states. He believes that they exist, however he is somewhat unique in that he claims to have shown how it is possible to maintain a belief in truly intentional mental states while staying within the realm of a physicalist ontology. In order to understand the problem for which Fodor claims to have a solution it is necessary to further investigate the two seemingly conflicting claims which his theory is supposed to reconcile. These claims are:

1. Mental processes are physical (syntactic).
2. Mental states are intentional (semantic).

In order to accommodate these two conflicting assertions Fodor needs to provide a naturalized, purely syntactic, set of conditions which are sufficient for a mental state to have an intentional content. His solution falls within a category called Causal Theories of Content and it solves some of the traditional problems of such theories by appeal to a criterion he calls Asymmetric Dependence. The discussion which follows aims to show that Fodor's theory does not succeed and that its failure is an indication that any effort to naturalize intentionality will also fail.

## 1. Why Mental Processes are Physical

Light strikes the surface of an object. Some of this light is absorbed, some reflected. Some of the reflected light travels through space and falls upon the rods and cones within my eyes. The incident of this electromagnetic radiation causes certain electrical charge signals to be sent along my optic nerve to my brain (the cause of these signals is an electro-chemical reaction). These signals reach my brain and here the picture gets a little muddy, if it hasn't been so far. What is safe to say is that electrical charges pass along neural pathways and chemical reactions occur as a result of the signal from my eyes.

At the end of all of that mess other electrical signals are sent out of the brain along the nerves which control muscle movement. My muscles twitch in accordance with the electrical signals they receive. The coordination of these muscle twitches is a behavior we know as shaking hands, performed on the object which the light had reflected off of (which we know as a woman). Physicalism, specifically its manifestation in neuroscience, demands that this is the whole story behind my action; there is nothing going on here which is not explained in terms of physics or chemistry.

These days pretty much everyone buys into the physicalism described above. Very few people want to argue that there is something other than neurons, blood, electrical charges, chemicals, etc. inside of our heads. Paul Churchland takes the general acceptance of physicalism a step further. He claims that any investigation or explanation of behavior must be done at this level. If the only thing going on inside of our heads is physical processes over the structure of our brains then our behavior must be explained in terms of these purely syntactical considerations. To put it differently, psychology should be neuroscience.



If we accept the dictates of physicalism and thus neuroscience then we must have a theory of mental processes which conforms to their dictates. The two major options which have arisen are Connectionism and the Computational Theory. Fodor adopts the latter of these as it implies a language of thought, an idea to which he is attracted. A computational theory claims that all mental processes are operations upon discrete syntactic (as opposed to semantic) structures. For Fodor, the syntactic entities upon which our mental processes operate are distinguishable neural firing patterns which symbolically constitute a representation of an external object. So there is a particular mental state composed of a neural firing pattern, certain chemicals, and other physical stuff which serves to represent, for example, cats. This mental state is instantiated whenever we have 'cat' thoughts and is thus the analog of a word in the human language of thought. Put simply, your mental processes are language-like operations on word-like (in that they are individually distinguishable) arrangements of neural pathways, electrical charges, and chemicals.

## 2. Why Mental States are Intentional

Consider the following: In a certain circumstance if you ask a friend walking on a sidewalk why she crossed from one side of the street to the other you are likely to get an answer resembling the following: "I wanted to avoid crossing the path of a black cat." If you happen to be a little slow on the inference side of reason you might in turn ask, "Well great, but what does that have to do with crossing the street?" A full explanation by your friend is likely to proceed as follows: "I saw a black cat cross

that sidewalk. I believed that a black cat crossing your path causes bad luck. I desired to avoid bad luck. I believed that I could do this by crossing the street and thus avoiding crossing the path of the black cat." Although most people are likely to reject the belief that crossing the path of a black cat causes bad luck, they are very prone to accept the kind of story which your friend told about why she crossed the street.

This type of story is an example of Belief/Desire Psychology, sometimes called Folk Psychology. Belief/Desire Psychology says that we have intentional mental states such as beliefs and desires which in general are the cause of our actions. This theory appears intuitively true to almost everyone upon only a little introspection. John Searle captures the seeming obviousness of folk psychology when he says, "it is hard to come right out and say that no one in the history of the world ever drank because she was thirsty or ate because he was hungry (p. 5)." Further it seems clear that the beliefs and desires which your friend claims to have as causes of her action are truly intentional.<sup>1</sup> That is, her belief that a black cat crossing your path brings bad luck is *about* cats (the black variety), directions of travel, crossing actions, and luck (the bad kind).

Now, if we do a survey of everything that exists in the universe we discover the uniqueness of this intentional content. Worms aren't about anything. Spiral nebulae don't have meaning in this way. Chairs don't have intentional content. But the thoughts which you just had seem to be *about* worms and *about* spiral nebulae, and *about* chairs. In fact, upon a

---

<sup>1</sup> What is not so obvious is the nature of this intentionality. Fodor makes two claims about the kind of intentionality which we are going to end up with. The first is that our intentionality is broad rather than narrow. This amounts to the fact that it is a representation of external (to our minds) objects, not internal objects (other mental states). This however has certainly not been a point of consistency in Fodor's writing.

The second distinction which Fodor makes is to adopt an informational version of intentional content. This theory, supported by Fred Dretske, is that "means" is synonymous with or reduces to "carries information about." Thus Fodor's theory is fully a Causal Theory of broad, informationally construed intentional content.

comprehensive survey the only things in the universe which seem to have this aboutness are our mental states and perhaps our language symbols by association. This uniqueness is certainly one of the defining characteristics of intentionality and must be explained in any theory of mental content.

### 3. The Big Problem

So it seems to me, and almost everyone else (philosophers included), that in the story related above, I shook hands with the woman because I **believed** that she was friendly and thus **wanted** to greet her, **believed** she would respond by doing the same, etc. It further seems that the wants and beliefs were *about* that girl in an important way, a way in which the kiss itself was not. Churchland, however, will disagree and certainly not retreat under the weight of mere public consensus. He will insist that all of this talk of beliefs and desires is delusion and will only mislead us in investigating behavior and the human mind. It is unclear, to say the least, how a certain neural firing pattern (or any merely syntactical object) could be *about* anything in the way which our mental states seem to be *about* things. And further, it is unclear, if this meaning is a purely physical phenomenon, why it is unique in all the physical universe to our mental states. In fact, intentionality appears to be incompatible with a naturalized, physicalist, theory of the mind. It looks as though either the scientifically well-justified physicalist theory of mental processes, or the intuitively obvious fact that we have intentional mental states which cause our behavior, must be abandoned.

All of this talk of mental states becomes complicated very quickly when we look at the complexity of the human brain. In order to simplify the discussion of this issue we talk in terms of the simplest mental state possible: a token of a single thought, such as 'cat'. If a theory which explains the intentional content of a single thought in purely physical terms is developed, then barring complications it may be assumed to apply to the full complexity of human mental states. In these terms then the problem of intentionality is reduced to the following question: "How can a single mental thought such as 'cat' manage to be about cats?" Fodor wants to show that this question can be answered in a way that does not violate the dictates of scientific physicalism.

#### 4. Fodor to the Rescue with the Big Problem

Fodor's solution to the big problem of reconciling intentionality and physicalism must provide conditions which are sufficient for a fully naturalized mental state (thus compatible with physicalism) to have an intentional content. The theory which his solution employs may be classified as a Causal Theory of Content (CTC).<sup>2</sup> The defining feature of this school of solutions is the claim that the content of a mental state is defined by reference to its causal relationships, what causes it and/or what it causes.

As Fodor's particular theory is one of broad content, it claims that the content of a mental state is defined in terms of the external objects which cause it to be tokened. Roughly speaking your thought

---

<sup>2</sup> Within this class I include both Behavioral and Functional labeled theories. Robert Cummins refers to these collectively as Covariationist Theories of Content. This seems to avoid any difficulty with implying a metaphysical claim about causation while still doing the work for which Fodor and his cohorts wanted it. Cummins' terminology will thus work for me in practice as well as in abbreviational convention.

'cat' means<sup>3</sup> *cat* because it is reliably caused by cats ( $\text{cat} \rightarrow \text{'cat'}$ ).<sup>4</sup> Notice that a CTC succeeds in remaining within the domain dictated by physicalism. It employs physical matter: mental states (which are composed of neurons, chemicals, electric charges, etc.), objects in the world (which are composed of carbon, oxygen, hydrogen, etc.), and a purely physical relationship between these objects, that of causation.

## 5. The Little Problem

An initial difficulty turns up for Fodor in the form of the Disjunction Problem. The Disjunction Problem arises from the inability to distinguish between true tokens of a disjunctive concept and false tokens of a singular concept. A standard formulation of the DP can be drawn in terms of the mental symbol 'cat'. While it is true that there is a causal covariance between cats and 'cat' it is also true that there is a causal covariance between rats and 'cat'. Specifically large rats sitting in the shadows may

---

<sup>3</sup> I will use "intentional content", "meaning", and "aboutness" interchangeably. Though this practice is certainly questionable the important thing for the argument is that all of these are interpreted as implying a semantic element.

<sup>4</sup> I will abide fairly closely to Fodor's notation but for the sake of precision my conventions are given below: " $\rightarrow$ " means causes (there seems to be some confusion as to whether this needs to be a law-like causal relation, but that debate will be ignored).

CAPITAL LETTERS distinguish mental concepts or types.

'single quotes' distinguish tokenings of mental concepts or types which could be either mental states or the corresponding language symbol which represents that mental state.

*italics* distinguish meanings and roughly it can be assumed that any word in italics should be preceded by the words "the property of being a" as Fodor's theory assigns properties as contents.

Therefore, fully, Fodor would say, "'Cat' is a tokening of the mental concept CAT which is about *cats* because it is caused by cats."

reliably cause 'cat' tokens. So, if it is true that "cats  $\rightarrow$  'cat'" and "rats  $\rightarrow$  'cat'" then, according to a CTC, 'cat' means the disjunct *cat or rat*. Further the Disjunction Problem implies that under a CTC any tokening of 'cat' to a rat must be a correct tokening. In fact there is no such thing as a false tokenings. Anything which causes a token of a symbol is part of the content of that symbol by the definition of a CTC. It seems that the content of our concept CAT is the singular *cat*, not the disjunct *cat or rat*, and that sometimes we do make errors in tokening 'cat', specifically in cases where we do so in response to a rat. If a CTC is to be taken seriously it must at least allow, if not provide an account of, this tokening error and either employ singular concepts or explain its reliance upon disjunct concepts.

If this wasn't bad enough Fodor voluntarily (and it seems correctly) makes matters worse for himself by noticing that the Disjunction Problem is merely a single outcropping of a more general problem which may be termed the Robustness of Meaning. "But surely this [teleological theories] underestimates what one might call the *robustness* of meaning; In actual fact, 'cow' tokens get caused in *all sorts* of ways, and they all mean *cow* for all of that. (1990 pp.90-91)" Here he is thinking of mooing noises, milk, manure, etc. Not only do other animals cause 'cow,' but many types of things do.

So Fodor has placed himself in a situation where he must argue that a mental state has its content in virtue of what causes it but that the content of mental states is independent of what causes them in an important way. Fodor's theory must provide a strictly naturalized account of the intentional content of mental states which allows for Robustness of Meaning and avoids the Disjunction Problem. He attempts to do just this in *A Theory of Content II*. His solution to the problem of robustness is a criterion he calls Asymmetric Dependence.

6. Fodor to the Rescue with the Little Problem

Asymmetric Dependence (AD) may be explicated in terms of the paradigmatic cats and rats example. According to AD the property of causing 'cat' tokens in rats is Asymmetrically Dependent upon the property of causing 'cat' tokens in cats. By AD Fodor means that if you broke the cat causes 'cat' relationship you would not have the rat causes 'cat' relationship but not visa versa. Fodor's complete formula for deriving content is then presented below.

'X' means *X* if:

- 1) "Xs  $\rightarrow$  'X's" is a law.
- 2) Some 'X's are actually caused by Xs.
- 3) For all Y not=X, if Ys qua Ys actually cause 'X's, then Ys causing 'X's is asymmetrically dependent on Xs causing 'X's.

One way in which Fodor formulates his method is to say that iff there are Nomologically Possible (NP)<sup>5</sup> worlds in which, "Xs  $\rightarrow$  'X's" is broken and so is "Ys  $\rightarrow$  'X's" but none in which "Ys  $\rightarrow$  'X's" is broken and so is "Xs  $\rightarrow$  'X's," then Ys causing 'X's is Assymmetrically Dependent upon Xs causing 'X's (X is the asymmetrically primary cause). In terms of the example above there must be NP worlds in which rats don't cause 'cat' but cats cause 'cat' but no NP worlds in which cats don't cause 'cat' and rats cause 'cat.'<sup>6</sup>

---

<sup>5</sup> The general idea of a NP world is that they are worlds in which the underlying principles of our psychology and physiology are the same but in which some nomic relations have changed.

<sup>6</sup> The talk of NP worlds is relevant because the causal connections involved are supposed to be counterfactual supporting. They are the result of the mental state's predisposition to be caused under certain circumstances.

## 7. Block's Obstruction

Ned Block presents an important clarification for AD which causes Fodor to modify the story he tells (1990 pp.111-114). Block observes that if Fodor is to succeed in providing a naturalistic account of content he must treat the criteria of his theory as simple physical relations which, if satisfied by objects, result in meaning. This means that 'X' is truly a variable which literally anything may satisfy. Specifically Fodor must interpret 'cat' as a phonological/ orthographic sequence with no semantic connotations. If 'cat' is viewed in this way, as merely a physical structure or pattern, then there is nothing to prevent there from being worlds in which rats cause 'cat' (or blenders cause 'cat' for that matter) and cats don't. Block argues that in order to claim the AD of "cats → 'cat'" Fodor must appeal to a preconceived semantic notion, that of 'cat' meaning *cat*.

It is important to keep in mind here that Fodor must provide a purely syntactical process by which a mental state may have a robust, non-disjunct intentional content. He must state his criteria in purely physical terms. It also behooves him to show that it is possible for his criteria to perform their function for at least a paradigmatic case. For his theory to be taken seriously he must not make an illicit appeal to semantic considerations. It is this last point which Block's objection attacks. He claims that AD presupposes semantic considerations as it is described in Fodor's theory.

Fodor recognizes that in order for his theory to be successful he must treat 'cat' as a phonological/orthographic sequence. However, simple revision of the AD story gives a theory which doesn't succumb to Block's objection. First, Fodor concedes that there would indeed be NP worlds in which "rats → 'cat'" holds but "cats → 'cat'" doesn't. However, he states that these NP worlds are



"farther away" than those in which "cats → 'cat'" holds and "rats → 'cat'" doesn't. In order to reach these worlds two changes must be made in the nomic relations instead of one for the worlds in which "cats → 'cat'" holds and "rats → 'cat'" doesn't. He explains this in the following passage:

"However, the asymmetric dependence proposal is that *all else being equal*, breaking cow causes "cow" breaks X causes "cow" for all X. Correspondingly-to put the point intuitively-what's wrong with Block's argument is that all else isn't equal in the worlds that he imagines. To get those worlds, you need to suppose *not only* that cow causes "cow" is broken, *but also and independently* that tree causes "cow" is in force. It's that independent supposition that violates the 'all else equal' clause." (1990 pp.112-113)

It seems that what Fodor has in mind is a sort of map of the NP worlds, with our world being the central one. There are then concentric circles of NP worlds, the first of which (the nearest possible worlds) is the result of changing one nomological relationship which exists in our world. For the purposes of the above example either "rats → 'cat'" or "cats → 'cat'" is false. Everything else in these counterfactual worlds is exactly the same as in our world. Except of course that if any nomological relationships were dependent upon the broken relationship then the dependent relationship will be broken too. If, when we investigate these nearest possible NP worlds, it turns out that the situation is as Fodor has described, then we have Asymmetric Dependence. So AD exists without appeal to semantic criteria and is thus acceptable within Fodor's physical/syntactic theory of content.

## 8. Fodor Overlooks Perception, a Revision to AD

In *Meaning and Mental Representation* Robert Cummins criticizes Fodor's Asymmetric Dependence. Cummins attempts to show that, in fact, AD never holds and is therefore unsuccessful at solving the problem of robustness. Although Cummins's argument is unsuccessful as a criticism of Fodor's theory it does correctly note that in assigning intentional contents Fodor's view of the causal relationships involved is too simple. The most obvious negligence of Fodor's account of the causal relationships involved in mental contents is that he has overlooked the importance of perception.

As described in section 7, Fodor's theory states that the counterfactual worlds relevant to AD are the NP worlds closest to ours, the ones where we break one causal relationship. For our purposes we break either "cats  $\rightarrow$  'cat'" or "rats  $\rightarrow$  'cat'." The causal relationships Fodor has in mind then are between a physical object (a cat) and a mental state (a 'cat' token). Though he does not discuss a mediating mechanism of this causal relation, a way in which it is manifested, it is safe to assume that Fodor would agree that there is one.<sup>7</sup> As he is trying to remain a physicalist, Fodor is likely to further agree that the mediating mechanism for these causal relations is perceptual.

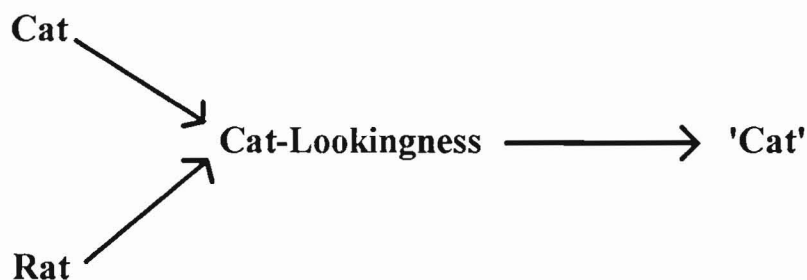
Consider the following account of the causal relationship between cats and 'cat's. I receive some proximal stimuli from a cat (reflected light, sound waves, odors, etc.), the sensory stimuli is transferred to the brain via neural firings along my sensory nerves, and (without getting **too** specific)

---

<sup>7</sup> Fodor does not want to argue for some sort of direct representation whereby the external object *just does* cause the mental state.

these neural firings cause the tokening of the mental state 'cat'. The array of proximal stimuli may be termed a Mediating Perceptual State (MPS).<sup>8</sup>

As Fodor wants to accept a naturalistic view of mental processes, it seems that he would gladly adopt this story, and in fact be hard pressed to deny it if he were so inclined. Cummins points out that given this account of mediating mechanisms there are two causal connections which Fodor's theory must accommodate.<sup>9</sup> The first is between cats or rats and the MPS, what he calls "cat-lookingness." The second is between the MPS and a tokening, 'cat,' of the mental state CAT. Thus a real story about AD must specify which of these is broken when investigating a NP world. The picture which Cummins has in mind is represented in Figure 2.



Cummins then claims that in order for AD to hold, the following two counterfactuals must have the truth values indicated:

1. If cats don't cause 'cat' then rats don't cause 'cat.' **True**

---

<sup>8</sup> It is okay for the exact distinctions between the object, the mediating mechanism, and the mental state to be somewhat vague. What is important for Cummins' objection is that there is indeed something "in-between" the cat and the 'cat'. Whatever this is; a pattern of electromagnetic radiation, a neural firing array, an electric signal, this is the mediating mechanism.

<sup>9</sup> Of course there are in fact quite a few causal connections involved here. What is important for Cummins argument though is that it is possible to assemble the causal connections into groups and give them the general names which he has specified.

2. If rats don't cause 'cat' then cats don't cause 'cat.' **False**

The heart of Cummins' attack is contained in the following passage. Reference to Figure 2 will help with the argument.<sup>10</sup>

Start with [counterfactual] (2). [There are] two ways to break the rat-to-|cat| connection: (i) Cat-ey looks don't cause |cat|s. But then cats won't cause |cat|s either. Thus, (ii) rats don't cause cat-ey looks. . . . But *cats* still look cat-ey, and hence they cause |cat|s. So (2) is false, as desired.

Now consider (1). Again [there are] two ways to attack the cat-to-|cat| connection: (iii) cat-ey looks don't cause |cat|s. Since, by hypothesis, rats poach by looking cat-ey, they will also cease to cause |cat|s, and (1) is true, as required. Unfortunately, this way of making (1) true makes (2) true as we just saw, so (iv) cats don't cause cat-ey looks. . . . But this won't affect the rat-to-|cat| connection, so (1) is false, contrary to requirements. . . .

The possible worlds in which (1) is true and (2) is false are not the same possible worlds. To put it another way, there is no single interpretation that makes (1) true and (2) false. Therefore, (1) and (2) do not jointly express something about the connection between cats and rats and that between cats and |cat|s. (1989 p.61)<sup>11</sup>

Cummins' claim amounts to the proposition expressed in the final paragraph, that the two counterfactuals required for Asymmetric Dependence never have the appropriate truth values in the same possible world. Notice, however, that this claim is fundamentally confused. These counterfactuals must have the appropriate truth values in the real world, not a counterfactual world. The counterfactual worlds are used to investigate the truth values of these counterfactuals in the real world, our world. Cummins is asking us to consider the status of these counterfactuals in counterfactual worlds. This of course necessitates investigating "second order" possible worlds which are accessible from the "first order" possible worlds which are counterfactual to the real world. Thus

---

<sup>10</sup> Cummins example is in terms of mice/shrews but I have edited the passage for our cat/rat example.

<sup>11</sup> Cummins uses absolute value marks to denote mental states and their tokenings.

he appeals to two degrees of modality. Fodor is only interested in the status of these counterfactuals in our world. This necessitates investigating only 'first order' counterfactual worlds and thus involves only one degree of modality. Cummins' argument thus fails to provide a basis for criticizing Fodor's theory as it confuses the modal structure upon which it is founded.

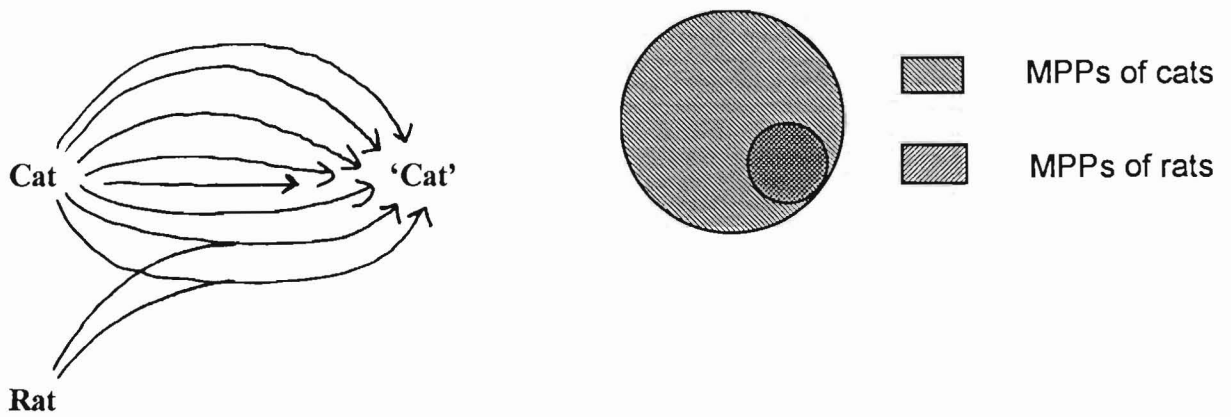
Cummins is, however, correct to notice the existence of a mediating mechanism in the causal relationships between objects and mental states. In fact he does not take this concept far enough. There are many causal relationships involved in each single causal situation. The molecular make-up of the cat causes the incident light to be "encoded," the encoded light causes the rods and cones to undergo certain electrochemical reactions which cause certain electrical signals to be sent along the optic nerves. Here the picture gets muddy because of the incompleteness of neuroscience, but suffice it to say that this neural array causes the mental state 'cat.' There are certainly many other causal steps within the brain but I will consider the point made.

It seems that Cummins was not taking his own point far enough in claiming the independent existence of a singular common intermediate state, cat-lookingness. For surely the causal chains by which 'cat's are ultimately caused are varied and distinguished at least up until the point where it gets muddy; after the neural impulses conveyed by sensory nerves enter the brain. But suppose that Cummins claims that there is a single neural array in which all of these causal chains eventuate, which in turn causes the mental state 'cat', and that this is "cat-lookingness". Wouldn't we then say that that singular mental state which is sufficient and necessary for 'cat' is in fact part of 'cat'? And if Cummins wishes to claim that it is not, hadn't he better have a good reason to do so? And isn't it terribly unlikely that he will have one given the "muddiness" of the brain and the obvious logic of incorporating it into

the mental state 'cat'. So, what Cummins calls Mediating Perceptual States are more accurately described as Mediating Perceptual Pathways (MPP) and a more precise representation of them is given below in Figure 3.

Now the thing to notice is that there are in fact many MPPs which fall under the category of cat-lookingness. There are cat-from-the-side-lookingness, cat-from-above-lookingness, cat-peeking-in-the-door-lookingness, cat-silhouette-lookingness, cat-on-its-back-lookingness, etc. etc. And this is not even to mention the other senses which include their own assortment of cat-smellingness, cat-feelingness, etc. So there is a huge set of MPPs which result in the mental token 'cat'. And as each of these corresponds to a unique causal chain, at least up until the point where it gets muddy in the brain, there are correspondingly many sensory neural arrays which result in 'cat'.

Once it is recognized that there is a set of MPPs which may lie "between" cats and 'cat's, a new interpretation of Asymmetric Dependence becomes available. What is necessary for Asymmetric Dependence to hold is that the MPPs by which rats cause 'cat' be a subset of the MPPs by which cats cause 'cat'. The NP worlds in which we are interested are those in which a subset of what we will call 'cat'-MPPs have been broken. So perhaps all of the cat-lookingness subset is broken; everyone becomes blind. It would still be true that "cats  $\rightarrow$  'cat'" by meowing but probably not that "rats  $\rightarrow$  'cat'." Thus, one causal relation is Asymmetrically Dependent upon another if the perceptual pathways by which it is mediated are a subset of those by which the latter causal relation is mediated. This situation is shown both in terms of the MPPs and in terms of the set of MPPs in Figure 3.



Under this version of AD Cummins's "cat-lookingness" becomes merely the subset of ways by which cats cause 'cat's that cats and rats share in common. Thus, in a world in which cats didn't cause "cat-lookingness" they would still cause 'cat' by other MPPs. What is important for AD is that cats cause 'cat' by a preponderance of MPPs, and that the MPPs by which rats cause 'cat' are a subset of these, and this does indeed seem to be the case. So rats poach off of the set of MPPs by which cats cause 'cat' as illustrated in Figure 3 above.

This version of Asymmetric Dependence seems quite amenable to Fodor's project. Instead of forcing him into talk of nearest possible nomologically possible worlds he deals with the perceptual pathways by which our mental tokens are caused. This version seems to provide a more accurate account of reality as noted above. It also appears to be a more "naturalized" version of AD in that its components are unquestionably physical; as opposed to the troublesome possible worlds.

The issue of the attainability of Asymmetric Dependence, however, is not yet laid to rest. Pat Manfredi and Donna Summerfield attempt to make an argument which acknowledges the "heterogeneity" of the causal relationship between cats and 'cat' but still attempts to show that AD will not work. Their effort is to no avail, however, because it turns out that. . .

9. Small Green Cows Pose No Threat to Fodor

In their article "Robustness Without Asymmetry: A Flaw in Fodor's Theory of Content," Pat Manfredi and Donna Summerfield attempt to show that Fodor's Asymmetric Dependence never holds, and is thus an unsuccessful method of explaining the robustness of meaning. Their attempt rests on showing that even in the paradigmatic case of cows and 'cow' AD does not hold, even though 'cow' certainly has a robust meaning.

Manfredi and Summerfield ask us to imagine a NP world in which cows have, through mutation, become small, green, and their manure has come to smell like sweet basil. They claim that in this world "horses → 'cow'" holds even though "cow → 'cow'" would not for a thinker who is unaware of the changes which cows have undergone. This shows that the map of counterfactual worlds is, in fact, never indicative of AD; there is always at least one NP world which throws a wrench in the gears.

Manfredi and Summerfield claim that their counterexample takes into account the heterogeneity of mediating mechanisms by which cows cause 'cow's. In their example they have changed everything about a cow which would initiate one of the MPPs, except perhaps ones dealing with DNA. Their objection fails, however, when the full revision of AD outlined above in consideration of the heterogeneity of MPPs is taken into account.

Despite Manfredi and Summerfield's feeling to the contrary, the world in which "horses → 'cow'" holds and "cow → 'cow'" doesn't is further away than that in which the opposite is true.<sup>12</sup>

---

<sup>12</sup> It is one of the weaknesses of Fodor's theory that if the world of small green cows ever actually came about then he seems committed to saying that Manfredi and Summerfield's thinker no longer has thoughts which are about cows even though it seems intuitive that he still could. In fact he probably would from time to time: "I wonder where all the cows went to?"



Notice that Manfredi and Summerfield's world is counterfactual in that properties of objects within it have changed. The "increment" by which distance from the real world is measured is then the number of properties which have changed. As the MPPs by which horses cause 'cow', and thus the properties by which they initiate the MPPs, are a subset of those by which cows cause 'cow' it is clear that more "property changes" must be enacted to reach a world in which only horses cause 'cow' than to reach a world in which only cows cause 'cow'. Thus the world of small green cows that don't moo and whose manure smells like sweet basil is further away than the world in which horses are afraid of the dark and so stay out of shadows.

#### 10. How Fodor Fails

In *A Theory of Content II* Fodor presents the following two quotes:

. . . symbols and mental states both have representational *content*. And nothing else does that belongs to the causal order: not rocks, or worms or trees or spiral nebulae. . . the main joint business of the philosophy of language and the philosophy of mind is the problem of representation. . . How can anything manage to be *about* anything; and why is it that only thoughts and symbols succeed? (Fodor 1987 p.xi)

Clouds *mean* rain. Spots of a certain kind *mean* measles. . . In all such cases there is a law-like or nomological regularity connecting one type of situation with another. Instances of these regularities are cases in which one situation means something or carries information about another: and, of course, in such cases there need be neither minds nor symbols used by minds.<sup>13</sup> (Israel 1987 p.3)

---

<sup>13</sup> While Israel's point is made in reference to informational accounts notice that it is really a problem of any naturalized CTC.

The problem raised here by David Israel is known as Pansemanticism. A theory which is Pansemantic assigns meaning to generic non-meaningful entities. CTCs have a problem with pansemanticism because of their reliance upon relationships between objects to define meaning. The generic non-meaningful world is just chock full of relationships between objects. So if meaning is defined in terms of these plentiful and diverse relations then it looks as if there will be a lot of this pansemantic meaning around, including that of clouds, spots, worms, and spiral nebulae. While it seems intuitive that our mental states have intentional content, it seems equally intuitive that spiral nebulae do not. As stated in section 2, it seems that intentionality is unique in all the universe to human mental states, and this uniqueness is one of its defining characteristics. Therefore, any theory of meaning must be able to distinguish things such as spiral nebulae and clouds from human mental states and prevent the former from being assigned intentional contents.

Fodor attempts to formulate the problem of pansemanticism for standard informational Causal Theories of Content like the approach supported by Fred Dretske:

If after all, meaning reduces (more or less) to reliable causal covariance, then . . . "smoke" means *smoke* and 'smoke means fire'. But this can't be right. If it were, then (since "carries information about" is transitive) it would follow that "smoke" means *fire*; which it doesn't. (1990 p.92-93)

However, Fodor thinks that the addition of his AD criterion solves the problem of pansemanticism which plagues traditional CTCs.

"Smoke" tokens carry information about fire (when they're caused by smoke that's caused by fire). But they don't *mean* fire because their dependence on fire is asymmetrically dependent on their dependence on smoke. (1990 p.93)

But notice that this misses the true point of the pansemantic objection. The claim which pansemanticism challenges Fodor to refute is that smoke means *fire*. The claim which he actually does refute is that 'smoke' means *fire*. Those single quotes make all the difference in the world. Pansemanticism argues that CTCs result in the unacceptable conclusion that lots of regular *stuff* (e.g. the physical object smoke) have meaning in the same way as our mental states. What Fodor has shown is that, if there is pansemanticism, then he can still solve the robustness problem of meaning. In fact, Fodor's AD also solves the problem of robustness for the meaning of smoke, the physical *stuff* and thus assigns meaning to a non-meaningful object. This can be shown by plugging the appropriate causal relationships into Fodor's three criteria as stated in section 6.

smoke means *fire* if:

- 1) "fire  $\rightarrow$  smoke" is a law.
- 2) Some smoke is actually caused by fire.
- 3) For all Y not=fire, if Ys qua Ys actually cause smoke, then Ys causing smoke is asymmetrically dependent on fire causing smoke.

The asymmetric dependency clause may be explicated by taking Y = cigarettes. It is true that "cigarettes  $\rightarrow$  smoke" but this is AD upon "fire  $\rightarrow$  smoke." Notice that if you break the "fire  $\rightarrow$  smoke" relationship then you have also broken the "cigarettes  $\rightarrow$  smoke" relationship but not *visa versa* (suppose all cigarettes get dumped in the ocean).<sup>14</sup>

A few more examples of pansemantic meaning resulting from the application of Fodor's theory will help to clarify its failure and ground the intuitive obviousness of the general problem raised by

---

<sup>14</sup> Note that according to Fodor's theory everything besides fire which reliably causes smoke must be AD upon fire causing smoke. It is.

pansemanticism for any CTC. The first example is given in a situation which I will borrow from Daniel Dennett's *Consciousness Explained*.<sup>15</sup> Consider the act of ducking when a brick is flying through the air at your head; let's call this a looming brick. Now it is certainly true that a looming brick causes ducking but it is also true that a bird flying overhead might cause ducking, specifically if its shadow resembled that of a looming brick. But notice that if bricks did not cause a ducking response (humans had really thick skulls which bricks merely bounced off of), then neither would a flying bird, the bird poaches off of the looming brick for its ducking causing ability. Also notice that the looming brick would still cause ducking actions even if the bird didn't (humans had 360° vision and so never confuse a flying bird with a looming brick). So according to Fodor a ducking **means** looming brick, but this is obviously not the case.

An example which falls further away from human action might be useful for those who may insist that ducking does mean looming brick. To this end we look at progesterone. Progesterone is a human hormone which causes suppression of ovulation during pregnancy. The question of whether it was designed to fulfill this function can be left for another day. However, it is undeniable that the teleological story about its coming to fulfill this function involves the fact that it interacted with the proper receptor in such a way as to cause the cessation of ovulation. Now for obvious reasons humans have become interested in the ability to stop ovulation on demand and to that end a synthetic molecule, norethynodrel, which mimics the ovulation interrupting function of progesterone has been developed. So here is the problem for Fodor: progesterone causes cessation of ovulation, but so does norethynodrel. If progesterone did not cause cessation of ovulation (testosterone did), then neither would norethynodrel. However, if norethynodrel did not cause cessation of ovulation (the differences

---

<sup>15</sup> Dennett uses this example to make a different point but it can be interpreted to apply to our discussion.

between it and progesterone resulted in its exclusion by the receptors), progesterone would continue to cause cessation. Norethynodrel poaches off of progesterone by being progesterone-resembling in the correct ways, but its ovulation cessation ability is AD upon that of progesterone. So the cessation of ovulation **means** progesterone.

The point to be gathered from these examples is that it seems likely that for any "relational" criteria similar to AD, there will either be (1) at least one case where meaning is attributed where we can not allow it (to smoke), or (2) the criteria will be too strict to be satisfied by the causal relationships of mental states.<sup>16</sup> The examples above suggest that this is indeed a reasonable assumption to make. Given the enormous complexity and variety of causal interactions in the generic non-meaningful world, I propose that the burden of proof lies upon the proponent of CTCs to show that it is possible to differentiate intentional states simply by their relational properties.

Another strategy available to Fodor is to take inspiration from the clarification of AD presented in Section 8 and appeal to the mediating mechanism which we associate with meaning. He could argue that meaning is had only when the causal relationships involved are mediated through perception. Notice that there are two problems with this approach. First, it would require justification as to why perceptual mediation distinguishes otherwise similar causal relationships into ones that result in meaning and ones that don't. It seems that in justifying this choice Fodor must either say, "Because it works if we assume that," or else reply with some, at root, semantic reason for choosing perception. Neither of these responses is satisfactory. However, there is a better reason to question this technique. Note that the restrictions of a naturalized theory would demand that Fodor recognize that "perception"

---

<sup>16</sup> It might be interesting to see if appeal to internal causal relations, and thus a narrow content, would allow for a solution to this problem. My intuition is that it wouldn't.

is just a general term for what is really a physical/chemical phenomenon. It looks like requiring the mediating mechanism to be perception would involve something like saying that the mediating mechanism must include the compression of air, or electromagnetic radiation, or neural firings, or some combination of these. This redoubles the need for justification as to why these physical mediating mechanism are special. In addition it seems that any sufficiently creative person could come up with a pansemantic counter example which qualifies as meaningful under any purely physical criteria which Fodor would place upon the mediating mechanism of intentional content.<sup>17</sup>

So it seems that Fodor will not be able to escape pansemanticism given the limitation of his tools of evasion. To remain a naturalized causal theorist of broad intentional content and a supporter of the computational theory of the mind which implies physicalism he must reconcile himself to the fact that smoke means *fire*, ducking means looming brick, among many other things. I, for one, would begin to question my theory when it started to tell me things like this. Therefore, we turn to . . .

## 11. The Implications of Fodor Failing

Before we really begin to investigate the failure within Fodor's theory it is necessary to make a clarification. If Fodor is to remain within the realm of naturalized theories of content he must argue that your thought 'cat' is about *cats* **because** it is a mental state which has a tendency to occur

---

<sup>17</sup> Here it seems that the connectionist might have more luck than Fodor devising some criteria for which a suitable pansemantic counterexample could not be created. However, I still feel that the burden of proof lies upon the connectionist.

due to cats. Intentionality must be causal covariance combined with AD; they can not be mere properties of an independently intentional state.

As was shown in Section 10, however, making intentionality a functional relationship does not distinguish meaningful causal relationships from non-meaningful. It does not capture the unique "intentionality" of our mental states which is implied by intentional Belief/Desire Psychology. And it doesn't do the job of reconciliation because it doesn't show how purely physical mental states can have the kind of content that it seems they do. It seems that relational properties are not **how** intentionality is distinguished; this is not where intentionality lies.

If the argument against an appeal to physical substance succeeds, then it would seem that intentionality is not distinguished by the physical stuff over which it operates either (neurons, electrochemical reactions, etc.). The question then becomes: What avenues are left to a physicalist? It seems that the answer might well be that there are none. If intentionality does not lie in a functional relationship, a physical substance, or a combination of the two, then it is unclear where else a physicalist might look for it.

So it seems that if we accept Fodor's challenge as laid out in *Psychosemantics*, "... the main joint business of the philosophy of language and the philosophy of mind is the problem of representation. . . . How can anything manage to be *about* anything; and why is it that only thoughts and symbols succeed?" (Fodor 1987 p.xi), we may need more tricks in our bag than those provided by a conservative physicalism.<sup>18</sup> Of course, we may choose to reject Fodor's challenge and the intentionality of our mental states with it. In either case we are offending what to many is a deeply held

---

<sup>18</sup> I'm not exactly sure how to define what I mean by "conservative physicalism" but perhaps it involves something like the possibility of a Searlian viewpoint.

sentiment; either the truth of physicalism or the existence of meaning. However, it seems that Fodor's failure serves as a rather convincing example of why any attempt to naturalize intentionality by appeal to causal relationships, and likely by any method open to physicalists will not succeed. So much the worse for naturalizing intentionality.



## **Part II: A Criticism of Daniel Dennett's Theory of Consciousness**

Daniel Dennett is another contemporary philosopher who has jumped on the naturalization bandwagon. While Fodor's project was the intentionality of our mental states, Dennett is concerned with what is probably an even more basic aspect of human existence, our consciousness. His tact is slightly different from Fodor's in that he employs actual experimental data gathered in psychology and neuroscience. From this scientific data he attempts to show that our present beliefs about the nature of consciousness are generally wrong and that it is in reality a purely physical phenomenon resulting from the non-centralized but highly complex system of parallel processing units in our brain.

Dennett assumes that the great work of naturalization has already been performed in the conclusive refutation of Cartesian Dualism at the hands of physics. His concern, however, is that even in the face of this undeniable refutation people still continue to think in terms of the "Cartesian Theater of the Mind," mostly because a plausible naturalized account of the phenomenon we call consciousness has not been provided. This then is Dennett's task: to provide a physical account of how we come to think that we have this phenomenon called consciousness and why it is in fact not what it seems to be.

## **1. Dualism Discarded**

Dennett spends very little time on his conclusive rejection of Cartesian Dualism, one seven page section to be exact. He states that, "This confrontation between 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" (1991 p.35). What he has in mind here is that the interaction between mind and brain which dualism seems to necessitate is in violation of fundamental laws of thermodynamics (specifically the conservation of energy) and dualism must be rejected. Thus Dennett feels confident in his, "apparently dogmatic rule that dualism is to be avoided at all costs" (1991 p.37). For Dennett the debate between physicalism and dualism has been resolved excepting the dualist developing an argument which answers the above conflict with physics. What remains is the belief in some aspects of consciousness which were suggested by dualism but which are unsupported under a physicalist ontology.

## **2. The Persistence of Our Confused Intuitions**

Even though Dennett feels that philosophers and scientists, if not society in general, have accepted the fact that dualism is dead, he argues that there is a persistent belief in the Cartesian Theater. The Cartesian Theater is the theoretical locus of experience. It is where the film of our perceptions is played before a central and ultimate experiencer, the central "I". The belief in such

a central experiencer is so deeply ingrained within human thinking that even those who have firmly rejected the idea of dualism, from which it originates, find it irresistible.

The concept in question is perhaps best exemplified by analogy to an observation platform in front of a train track. The motion of the train represents the passing of time and each car represents a bundle of experience which passes in front of the platform. The central "I" sits on the platform and the order in which the cars pass by is the order in which they are perceived. The senses accumulate information and process it and then when it is fully processed and packaged it is "sent up" to consciousness where it is observed. This idea originates in the stream of conscious experience which seems to make up our lives, the fact that each moment follows the one before it and we seem to constantly be aware of what is happening, if to a greater or lesser extent. Under a dualist ontology the mind was the central observer for which the brain put on a show in the Cartesian theater. However, under a physicalist ontology there is no such candidate. It appears that there is no central locus where all the information in the brain is ultimately available to the central "I" which watches the Cartesian Theater. So Dennett's radical claim is that the stream of consciousness which seems to constitute our lives is a grand illusion and in fact, there is no central perceiver. There is no film of experience because there is no theater in which to watch it.

### 3. What Consciousness Really Is, the Multiple Drafts Model

Dennett maintains that our brain is composed of a massive array of independently operating parallel processors, many of which are specific to certain tasks. These processors supply information to various other parts of the brain in order that they can perform their function. The entire system is set up to work together and deal with the world in such a way that the human organism can attain the things which it needs. The purpose of the brain, as Dennett puts it, is to predict future, to determine what will happen next so that the organism may act appropriately. To this end it must operate quickly, providing information at the earliest possible time. However, it must also operate accurately, revising the information provided as new data is gathered and further processing occurs. Thus Dennett postulates a brain that handles information in a way which is analogous to documents within an electronic mail system. The information provided by the different processors is like a letter sent out in draft form to any department to which it is relevant. In light of further processing new drafts will be sent out and if different departments make revisions they will send out new drafts as well. At any given time action is performed based on the current draft available to a department. Of course ultimately revisions cease to be made and the information is filed, or stored in memory. Until the piece of information is settled in memory it is in a state of perpetual revision. Thus, there is no single draft which may be labeled as the one had by consciousness. There is no continuous flow of single experience from moment to moment but merely available information at any given moment. This theory of how the brain operates he calls the Multiple Drafts Model.

In order to justify the adoption of his Multiple Drafts Model of consciousness it seems that Dennett needs to accomplish two tasks. He needs to show how our brain could have come to function in such a way that it seems to us that we have a stream of consciousness and yet do not. To this end Dennett presents a full account which attempts to explain the evolution of what we call consciousness, and how a belief in the phenomenon humans claim to experience would be selected by evolution. That aspect of Dennett's theory will not be discussed in this paper. Dennett's second task is to provide an impetus for accepting his view of how the brain works in the face of our intuitions. Here what is required is an argument against the intuitively obvious view that we do indeed have immediate awareness of our experience and that this awareness is a stream-like, continuous, time ordered sequence. Dennett's claim is that this apparent phenomenon falls apart on a small time scale. The argument which he gives for this is in terms of psychological experiments performed on what is called the color phi phenomenon.

#### **4. Evidence Against a Central "I," the Color Phi Phenomenon**

The color phi phenomenon provides a clear case where the distinctness of Dennett's Multiple Drafts Model can be seen. The experiments on the color phi which Dennett employs were performed by Paul Kolers and Michael von Grunau in 1976. The experiments consist of an observer watching a screen where two dots of light, one red and one green, which are separated by as much as 4 degrees of visual angle, are alternately flashed for periods of 150 ms with a 50 ms

interval between the two. The observer, however, reports seeing a single dot which moved from the first location to the second and changed color halfway between the two.

The observer is obviously performing a revision upon the information which their senses provide. There were in fact no intermediate “frames” in which the light moved and changed color. According to traditional versions of consciousness they are “adding in” the frames of experience which constitute the light moving and changing color. Dennett distinguishes two types of revision which may occur according to a proponent of traditional consciousness. The first he calls Stalinesque Revision because it involves the staging of a false reality. A Stalinesque Revision would occur before consciousness; it involves the appropriate processor in the brain judging that the best interpretation of the data given is that the dot moved from left to right and changed color in the middle. This or another processor then creates the “missing” frames of experience which are deemed to occur in between the ones actually perceived. These frames are then inserted into the information stream which is subsequently sent up to consciousness for viewing. The alternative is Orwellian Revision because it involves the rewriting of history. An Orwellian Revision occurs after consciousness but before the observer has the opportunity to respond or even form a permanent memory of conscious experience. The story here is that the person really is conscious of the two separate dots but then a processor in the brain notices that a better interpretation would be that of a moving, changing dot and thus edits the memory of the experience to this end, inserting the “missing” frames. By the time the person is able to respond the memory of the event has been edited so that what they actually remember and report is not

what they experienced in the moment. So Stalinesque Revisions are correctly reported false experiences and Orwellian Revisions are falsely reported correct experiences.

Dennett here wants to present an argument which shows that while these may seem like valid alternatives on the surface, in fact there is no distinction between them. He notes that both theories predict the same behavior and responses. In fact both theories claim that the experience will “feel” the same to the observer, at least in retrospect. So not only is it not possible to perform a test which determines which is the correct theory, it is not possible for a person to determine which process is occurring in their own brain.

The two theories tell exactly the same story except for where they place a mythical Great Divide, a point in time (and hence a place in space) whose fine-grained location is nothing that subjects can help them locate, and whose location is also neutral with regard to all other features of their theories. This is a difference that makes no difference. (1991 p.125)

From this Dennett presents what appears<sup>19</sup> to be the following argument:

- 1) If there was a Cartesian Theater, a point of consciousness, then there would be real difference between the Orwellian and Stalinesque theories of revision.
- 2) As it is impossible in any way to distinguish between the theories, there is in reality no difference between them (verificationism).
- 3) Therefore, there is no point of consciousness, no Cartesian Theater. Thus there is no film to be shown in the theater, and no central “I” to watch it.

---

<sup>19</sup> Dennett’s argument is far from clear and is fraught with dogmatic appeals to the rejection of dualism but I believe that this is an accurate presentation of the underlying points being made.

Dennett is aware that verificationist, or operationalist, arguments of the type presented in premise 2 are rejected off hand by many people and usually contentious at best. However, he argues that in this case it is appropriate to make a verificationist claim, as what we are in fact talking about is the subjective experience of the observer. To deny premise 2, Dennett claims, is to posit what he calls the objectively subjective, “the way things seem to you even if they don’t seem that way to you” (1991 p.132). The Stalinesque and Orwellian theories of revision each posit a different version of immediate experience; Stalinesque Revision says that the immediate experience was of a moving dot and Orwellian Revision says that it was of two distinct dots. However, as noted above, the observer can not tell which one of these immediate experiences occurred. To claim then that there was in fact a difference in experience which can not be distinguished is to claim that the experiencer has imperfect access to their immediate conscious experience which certainly seems wrong and very likely is paradoxical.

Dennett concludes by stating his radical claim based on this argument that,

There is no reality of conscious experience independent of the effects of various vehicles of content on subsequent action (and hence, of course, memory). . . . When you discard Cartesian dualism, you really must discard the show that would have gone on in the Cartesian Theater and the audience as well, for neither the show nor the audience is to be found in the brain, and the brain is the only real place there is to look for them. (1991 pp.132,134)



## 5. Two Problems With Dennett's Theory

The first and perhaps more pervasive difficulty with the theory of consciousness proposed by Dennett is that it seems to have missed the central fact about human consciousness, immediate experience. This is the notion that there is more than mere information processing going on in our heads; there is experience. You can relax your eyes and just see the basic outlines of writing on this page, or you can focus in on the words and decipher them but you can not stop seeing the page if it is in your visual field. When you bite into a lemon it seems sour to you. You don't just believe that it is sour but it actually tastes sour. Your stream of experiences is continuous, excepting sleep, it never stops; your visual field is always "in front of you." It is hard to capture in words what is obvious to almost everyone; humans have unavoidable, immediate, continuous and undeniable experiences which go beyond the processing of information.

Dennett never truly addresses this issue. He continually refers to experience, even under his new theory, while not mentioning the fact that by this word he means something very different from the phenomenon which most people believe themselves to have. He does not provide an account of how we come to fool ourselves into our belief in immediate experience, though he skirts the issue several times. As far as I know no physicalist interpretation of experience has ever been very satisfying to a realist about experience and it doesn't seem as if Dennett's particular version provides any new insights into this entrenched difficulty. While Dennett's avoidance of this problem and seeming inability to deal with it is the most serious flaw of his theory it is also the most difficult objection to make. There is not much to be said here except that it seems as if

there is more than what Dennett's theory gives us. Dennett would likely choose to insist that the "experience" which his theory allows for is the experience we have. If he asks us to clarify what we think is lacking in the experience he allows for we will be hard pressed to come up with something tangible and credible. This, however, is not the only ground upon which Dennett's arguments may be attacked. There is also a much more concrete avenue by which to object to his theory.

It turns out that upon closer examination the argument which Dennett presents in an attempt to justify his verificationist second premise is misconstrued. He claims that to say that there could be aspects of immediate experience of which the experiencer is unaware is to deny the concept of incorrigible access to immediate experience, which would seem to be a problem. However, the problem here is not that the experiencer need have missed or confused the **immediate** experience but that they are unable to distinguish a difference in retrospect. The time frame in which Dennett's example exists is hundreds of milliseconds. By the time an experiencer responds regarding any difference the experience they are basing their response on is already in memory and permanently altered under either theory of revision. So the inability of an experiencer to distinguish between the two theories of revision is indicative of human's imperfect access to memory, something which everyone is all too ready to accept. Again, the fact that an experiencer is unable to provide evidence, even to themselves, which would distinguish the two theories of revision, shows only that what is in our memories is not reliably exactly what happened in our immediate experience.

With this point taken it seems that Dennett's verificationist argument is subject to standard objections against verificationism and is likely to be rejected by most people. It just seems false to conclude that merely because we can not tell that there is a difference there really isn't one. And our intuition that we do indeed have an immediate stream of continuous, unavoidable experience which necessitates there being a real difference between Orwellian and Stalinesque Revisions will probably be strong enough that we will accept the existence of a difference even in the absence of evidence. As it turns out we will likely never be able to tell which revision theory is correct. In fact perhaps they are both employed in different cases, but our inability to determine the actual nature of our brain's operations is nothing new.

## **5. Where This Leaves Dennett**

It is unclear what the effect of this criticism is upon Dennett's entire project. As his argumentative structure is vague if not absent it is not easy to perform a damage assessment. However, the failure of his verificationist premise at least shows that Dennett's claim that there is no central locus of experience is not one which we must adopt. Dennett would certainly argue that it is still the best option given his dogmatic rejection of dualism. This, however, is certainly unsatisfying, especially if one is not yet convinced that physicalism, at least of the form employed by Dennett, must necessarily be adopted. While these arguments do not provide a knock-down.

objection to Dennett's theory, it does seem that they undermine the motivation which Dennett had provided for its necessity.

What is more, it seems that in light of the admittedly sketchy comments regarding Dennett's treatment of immediate experience, our view of his entire project should be somewhat shifted. The two options available to a physicalist in dealing with these "mental phenomena," remember, are to naturalize the phenomena or to claim that it does not really exist, at least as it is traditionally thought of. It appears that Dennett's theory results in the disintegration of immediate experience into mere information available for behavior modification. This is in effect to deny the reality of what consciousness is constituted of and thus classifies Dennett as not a naturalizer, but an anti-realist regarding consciousness. It also suggests that his book is mistitled and in fact should be called *Consciousness Explained Away*.

## Conclusion

So it seems that neither Fodor's attempt at naturalizing intentionality nor Dennett's theory of consciousness really deliver the explanatory goods. Fodor, it seems simply fails to provide an adequate account of the meaning of our mental states and Dennett's theory fails to be convincing and explains a phenomenon which leaves out the central component of our conscious life. What are we to think in light of these observations?

From the arguments presented in Section I it would seem that physicalism will not be able to cope with a realist view of meaning. It appears then that our options are to either reject the intentionality of our mental states or to adopt an explanation outside the realm of physicalism. To reject mental content, however, is to deny an immense and continuing tradition of belief in the intentionality of our mental states. It also seems, at least to me and I will wager to you as well, that it denies the fundamental reality of our mental lives, my thought 'cat' really is about cats. The situation is similar in the case of consciousness. The arguments presented above do not necessitate such a decision between rejecting consciousness and adopting an explanation beyond physicalism. However, the theory which Dennett has advocated does seem to present a version of "consciousness" which compromises much of what seems to be going on in our heads and approaches a form of anti-realism regarding consciousness. Here again then we are presented the option of denying a vast tradition of belief in a phenomenon which seems very alive within at least my personal experience.

If then we choose to maintain our intuitions regarding the content of our mental states we will need to posit something beyond the realm of physicalism as it is currently understood. This does not necessarily mean a return to Cartesian Dualism but certainly suggests a loosening of current ontological attitudes. Part of this loosening may be the realization that our current definition of physical substance is too naive and is in fact equivalent to: “what we can explain right now.” Once we have done this, it is a small step to further allow for a non-physical explanation of consciousness which would then preserve our intuitions regarding immediate experience. The arguments presented in this paper have shown that maintaining a strict form of physicalism requires denying basic facts of human existence (the intentionality of our mental states and consciousness as we know it). In light of this conclusion it may in fact be reasonable to posit an basis for these phenomena which we are not yet, and possibly never will be, able to explain by scientific investigation and in scientific terminology.

## References

- Cummins, Robert, 1989. *Meaning and Mental Representation*. Cambridge, MA: MIT Press.
- Dennett, Daniel, 1991. *Consciousness Explained*. New York :Little, Brown, and Company.
- Dretske, Fred, 1981. *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press.
- Fodor, Jerry, 1975. *The Language of Thought*. Cambridge, MA: Harvard University Press.
- Fodor, Jerry, 1987. *Psychosemantics; The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press.
- Fodor, Jerry, 1990. *A Theory of Content and Other Essays*. Cambridge, MA: MIT Press.
- Fodor, Jerry, 1994. *The Elm and the Expert; Mentalese and its Semantics*. Cambridge, MA: MIT Press.
- Havel, Václav, 1986. *Thriller*, in J. Vladislav, ed., *Vaclav Havel: Living in Truth*. Boston: Faber and Faber.
- Israel, David, 1987. *The Role of Propositional Objects of Belief in Action*, CSLI Monograph Report No. CSLI-87-72. Palo Alto: Stanford University.
- Manfredi, Pat, and Donna Summerfield, 1992. "Robustness Without Asymmetry: A Flaw in Fodor's Theory of Content," *Philosophical Studies*, 66, 261-283.
- Putnam, Hilary, 1975. *Mind, Language and Reality; Philosophical Papers, Volume 2*. Cambridge: Cambridge University Press.