INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.

2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of "sectioning" the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.
FUNCTIONALISM:
MATERIALIST THEORY OF MIND
OR
MENTALIST THEORY OF BRAIN?

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By
James S. Kelly, B.A., M.A.

* * * * *

The Ohio State University
1983

Reading Committee: Approved By
William G. Lycan
Alan Hausman
Robert G. Turnbull

William G. Lycan
Adviser
Department of Philosophy
TO

WARREN STEINKRAUS
(who started me)

ROBERT ALMEDER
(who encouraged me)

MY MOTHER
(who doesn't understand me)

JOIE
(who loved me)

THANKS
VITA


1971. . . . . . . . .  B.A., State University of New York
                  at Oswego, Oswego, New York

1976-1982 . . . . .  Teaching Associate, The Ohio State
                  University, Columbus, Ohio

1979. . . . . . . .  M.A., The Ohio State University,
                  Columbus, Ohio

1983. . . . . . . .  Visiting Instructor, Miami University,
                  Oxford, Ohio

FIELDS OF STUDY

Major Fields: Philosophy of Mind, Functionalism,
              Philosophical Psychology, Metaphysics

Areas of Competence: Ethical Theory, Legal Philosophy,
                      Philosophy of Language, Epistemology,
                      Logic, 17th Century Cartesian Philosophy,
                      History of Modern Philosophy
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEDICATION</td>
<td>ii</td>
</tr>
<tr>
<td>VITA</td>
<td>iii</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Chapter</td>
<td></td>
</tr>
<tr>
<td>1. PHENOMENAL PROPERTIES: THE PROBLEM</td>
<td>4</td>
</tr>
<tr>
<td>Problem Properties</td>
<td>5</td>
</tr>
<tr>
<td>Pro-Qualia Intuitions</td>
<td>9</td>
</tr>
<tr>
<td>Evolution and Causation</td>
<td>15</td>
</tr>
<tr>
<td>Emergence/Supervenience</td>
<td>24</td>
</tr>
<tr>
<td>Token-Identity: Types and Tokens</td>
<td>35</td>
</tr>
<tr>
<td>Neutralizing Introspection</td>
<td>43</td>
</tr>
<tr>
<td>2. INVERTED QUALIA</td>
<td>53</td>
</tr>
<tr>
<td>Seeing Red What is Green</td>
<td>54</td>
</tr>
<tr>
<td>Onland Churchland Science</td>
<td>57</td>
</tr>
<tr>
<td>Science and Introspection</td>
<td>86</td>
</tr>
<tr>
<td>3. ON LEAVING NOTHING OUT--THE FATA MORGANA TACK</td>
<td>95</td>
</tr>
<tr>
<td>Playing Our Game</td>
<td>95</td>
</tr>
<tr>
<td>Arguing for Nothing</td>
<td>98</td>
</tr>
<tr>
<td>Posits: For the Purpose of</td>
<td>114</td>
</tr>
<tr>
<td>4. PUTTING QUALIA ON THE LEVEL</td>
<td>123</td>
</tr>
<tr>
<td>Homunctionalism</td>
<td>123</td>
</tr>
<tr>
<td>Teleology (An Aside)</td>
<td>127</td>
</tr>
<tr>
<td>Pain as Analysable</td>
<td>130</td>
</tr>
<tr>
<td>Supervenience</td>
<td>138</td>
</tr>
<tr>
<td>Recurring Problems</td>
<td>160</td>
</tr>
<tr>
<td>Inverted Qualia--Resolution of a Sort</td>
<td>167</td>
</tr>
<tr>
<td>5. THE ROLE OF EXPERIENCE: PART I</td>
<td>174</td>
</tr>
<tr>
<td>Functionalism, Token-Identity, The Role of Experience--An Introduction</td>
<td>174</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Introspection as Brain Scanning</td>
<td>184</td>
</tr>
<tr>
<td>Private Language</td>
<td>188</td>
</tr>
<tr>
<td><strong>6. THE ROLE OF EXPERIENCE: PART II</strong></td>
<td>225</td>
</tr>
<tr>
<td>On Leaving Something Out</td>
<td>225</td>
</tr>
<tr>
<td>Attempted Solutions</td>
<td>234</td>
</tr>
<tr>
<td>Quasi-Epiphennomenalism</td>
<td>256</td>
</tr>
<tr>
<td>Appearances--A Hole in the Functionalist Net</td>
<td>267</td>
</tr>
<tr>
<td>Conclusion</td>
<td>285</td>
</tr>
<tr>
<td><strong>NOTES</strong></td>
<td>287</td>
</tr>
<tr>
<td><strong>BIBLIOGRAPHY</strong></td>
<td>315</td>
</tr>
</tbody>
</table>
INTRODUCTION

Some readers may think that my tone of voice in what follows is excessively polemical. It may comfort them to know that the assumptions against which I exhibit most heat are assumptions of which I myself have been a victim. Primarily I am trying to get some disorders out of my own system. Only secondarily do I hope to help other theorists to recognize our malady and to benefit from my medicine.

The above paragraph, of course, is an unabashed rewrite from Gilbert Ryle's introduction to The Concept of Mind. What I must impress upon the reader is that I have no "knockdown" argument against Materialism. Indeed, I am strongly motivated by naturalistic intuitions and desires. Yet it is motivation that is made of such stuff, not arguments--it is the latter I find the Materialist lacking. In a general way I hope to accomplish against the Materialist what many take Ryle to have sought to do against Descartes. By pressing hard on certain home truths, I seek to evaluate the standard materialist approach of the last twenty-five years. My arguments are not formally distinct from Saul Kripke's or Thomas Nagel's. Nonetheless, I find Kripke's employment of certain analytic tools in his cause has only succeeded in clouding the force of the non-materialist's
well motivated worries. Nagel, on the other hand, has not said enough.

Naturalism is a strong motivation for seeking to render Materialism as the correct account of a unified Philosophy of Man. Still, what I take to be standard materiasm, basically a third-person, topic-neutral account, is overly strained and does not offer an argument adequate for supporting Naturalism. Vis-à-vis this standard materialism, my view is non-traditional and not at all "trendy"; what I have sought and failed to find is a plausible argument, one that accommodates home truths, for Naturalism. At the risk of being cute, I see myself as challenging the dogma of the Machine in the Machine. My success, as William Lycan sees it, will put the ball back into the Materialist's court: just what is their argument for Naturalism?

When I have pressed my intuitively based disquiet concerning this standard materialism, I have been met with two typical responses. The first has been a somewhat obscure reference to the Private Language Argument concatenated with a topic-neutral reference fixing account. I have sought and failed to see how that argument makes Materialism plausible; its motivation is clear enough. My investigation of that argument begins in Chapter 1 and is most directly examined in Chapter 5. Another typical response, a response which leads me to believe the third-person, topic-neutral
reference tradition of the last twenty-five years is a tradi-
tion gone wild, has been in the form of a question: what
more do you want—what is the data that is not accounted for?
Again, Chapter 1 begins an answer to that question which is
pursued in the rest of what follows. I hope to make clear
what more is wanted by a Materialist account by bringing to
the surface the data the Materialist (as I view it) ignores.
Chapter 2 brings out the force of the intuition that drives
my disquiet, and Chapter 3 is an examination and refutation
of Daniel Dennett's attempt to squelch that intuition. In
Chapter 4, an explication of the most plausible form of
materialist functionalism is likewise found wanting.
Finally, Chapters 5 and 6 put it all together and issue a
demand that the Materialist face the hard questions, ques-
tions I maintain have been skirted since Place and Smart
first pumped new blood into the materialist cause. It is
important that the reader not take me as presenting a
methodological criticism about science; I applaud scientific
methodology—nonetheless certain data remain unexplained.
1. PHENOMENAL PROPERTIES: THE PROBLEM

For up to this point I have described the earth, and all the visible world, as if it were simply a machine in which there was nothing to consider but [the] figure and movements [of its parts], and yet our senses cause other things to be presented to us, such as colours, smells, sounds, and other such things, of which, if I did not speak, it might be thought that I had omitted the main part of the explanation of the objects of nature.*

What concerned Descartes troubles many of us today; we are still struggling to explain (or explain away) the presence of these "colours, smells, sounds, and other such things" whose qualitative phenomenal aspects, presented to us, are as often sought as (with luck) avoided--love and fine wine beckon, pain and the stink of rotten cabbage repel. Almost all of us are familiar with various sensations such as these and it at least seems to us (with a bit of introspective reflection) that some of our sensations have intrinsic non-relational qualities--some tastes, for instance, are sweet. Any theory claiming to solve the mind-body problem (or set of problems) and adequately explain human experience owes a heavy debt (witness Descartes' ontological/causal difficulties) unless these aspects of our common experience, these "objects" of nature, find a comfortable niche. To find a
comfortable niche within a theory of mind is to explain (be an explainer) or be explained (be an explainee). Pain, for example, may serve to explain why it is we avoid certain objects (red-hot irons) or situations (rock concerts). Of course pain may itself be explained in terms of neurons and chemicals. Whether pain is an explainer or an explainee may well determine the shape of a particular theory of mind.

How pains and other qualitative features of our phenomenal experiences are to be accommodated by (materialist) functionalist theories of mind is the present concern; for pain is not only a state that makes us flee or a state brought about by injury, but a state it feels like something to be in. First I shall set the general problem, leaving details until later.

Problem Properties

Descartes' 17th century attempt to secure a niche for both the explanatory scheme of the "new" science, which viewed the universe as an elaborate mechanism, and the older religious tradition and its emphasis on explanation in terms of God's purpose, spawned current interest in the mind-body problem. By arguing for the existence of a non-material substantive mind with a sui generis ontic status, Descartes sundered the world's stuff, thereby raising the spectre that a completely naturalistic account of the universe and its
denizens is beyond the pale of scientific theorizing (done in terms of purely physical principles).

Although this Cartesian dualism of substances is currently in ill repute (for cause), what we may call Descartes' "data base" still poses serious difficulties for materialist theories of mind. We can (I do) view his arguments for dualism as revolving around two features of the human condition: (i) we are directly aware via introspection, as presented to us in experience, of the existence of something ("objects of nature") which science has yet to explain adequately—here arises the problem of consciousness, and (ii) the explanation of the human capacity to adapt to changing circumstances by appropriately altering behavior, demands appeal to properties radically unlike the properties of matter—here appeal is made to goal directed behavior intelligently performed. Here, of course, I am reading into Descartes, but it is not, for that, completely off the mark. For support of (i) I here lean on the opening quote, and a plausible extension of it. With regard to (ii), consider Descartes' seeking of that special aspect of human experience which distinguishes Man from brutes or machines and which leads to his two tests:

The first is, that they could never use speech or other signs as we do when placing our thoughts on record for the benefit of others... it never happens that it [the brute] changes its speech in various
ways, in order to reply appropriately to everything that may be said in its presence, as even the lowest type of man can do. And the second difference is, that although machines can perform certain things as well as or perhaps better than any of us can do, they infallibly fall short in others... [thus acting]... only from the disposition of their organs. For while reason is a universal instrument which can serve for all contingencies, these organs have need of some special adaptation for every particular action.²

(emphasis mine)

I had described after this the rational soul and shown that it could not be in any way derived from the power of matter. . . .³

Here, as I interpret it in (ii) above, reason, the "universal instrument" allowing for appropriate language use and action, has properties radically unlike matter. I take it that (i) and (ii) are sufficient for claiming that dualism revolves around these features of the human condition, even for Descartes. For the classical Cartesian, res extensa is opposed to res cogitans; the distinction between body and mind is the distinction between that which is corporeal and that which thinks. Since, as he understood it, the word "thought" applies to "all that of which we are conscious as operating in us;"⁴ Descartes maintained that understanding, willing, imagining, feeling, etc., all fall under the rubric of "thought." Contemporary philosophical
views, however, have gone on to mark a further deeper dis­
tinction.

Persons have propositional attitudes, for instance, the
ability to believe that a Friday night rendezvous will be
rewarding, as well as the capacity to experience phenomenal
qualities, for instance, the felt pain of a match burn. Our
ruminating about an impending liaison and our feeling the
searing pain of a burn are both paradigm cases of mental
states. Nonetheless, we distinguish between these types of
mental states by focusing on their distinctively different
properties. To capture this distinction we need only recog­
nize the difference between thinking about a non-occurrent
pain and feeling a current pain. (Here we depart, it appears,
from Descartes.) In the former case we talk of the mental
state's intentional properties, while in the latter talk is
of phenomenal properties. Although both sorts of properties
pose (at least prima facie) problems for any materialist
theory of mind, I shall limit discussion to those phenomenal
features which have come to be called qualia (immediate
phenomenological properties). Though no satisfactory defi-
nition of "qualia" comes quickly to mind, the familiar
examples of perceived colors, pains, tickles, sounds, aromas,
tastes and rushes adequately marks the elements of concern
(unless one has been forover anesthetized). That these
elements do present problems stems from what we may dub the
"pro-qualia intuition."
Pro-Qualia Intuitions

Not only do we believe (preanalytically) that there is more to pain that its causal role, we believe that our immediate awareness of the feel of pain is awareness of that feature by which we distinguish (in our own case) a pain from a tickle or a searing pain from a throbbing one. What I call the pro-qualia intuition is the set of beliefs stemming from our introspective abilities and centered on the "what it is like" locution. Thus the belief that the features which endows the introspector with the ability to respond differentially to his inner states is that of which he is directly aware, namely, the feel. It is the "what it is like" feature of, for instance, being in pain that is alleged to be the stumbling block for materialism. (Note, I am not worried about the locution "what it is like to be," as in "what it is like to be a bat." Since pain at t is that particular experienced state which is the effect of a certain range of stimuli and the cause (it seems, see below, ch. 4) of certain responses, it would appear that "pain" is verbally indefinable—a word whose referent we can only learn ostensively via experience, albeit with the aid of stage-setting public circumstances (see below ch. 5). Consider a person who is congenitally unable to feel pain. Such a person cannot know what pain is because he does not know what it is like to be in pain, that is, does not know
what pain feels like. Though science may be able to explain why a person cannot feel pain by telling us about the person's causal shortcomings (atypical "wiring"), it is difficult to fathom how science can tell us something interesting about what it is that is at the end of the causal sequence of physiological items that leads to felt pain. Let us consider a thought experiment in order to expand upon this intuition (which will be investigated in detail in ch. 4).

Imagine a situation in which all persons with the exception of Duncan are subjected to radiation of a sort which so alters their internal workings that they are unable to feel pain as we are now able to feel pain. Furthermore, suppose this change to be genetic so that only Duncan's progeny are able to feel pain. As such they will be unto themselves a community of pain feelers such that only they will understand (know) that to which "pain" refers. Some five hundred years after Duncan's death his only living descendant is Bat. It seems clear that it makes perfectly good sense to claim there is something it is like for Bat to be in pain, but no one else knows what it is like—despite there being plenty of written descriptions by past pain feelers and as much physiological knowledge as you please. The scientists of the day, we can imagine, know know all the facts there are to know about Bat's physiology
and functional organization; they would not, however, know what it is like for Bat to be in pain. Nor could they come to know this short of experiencing a feel like that which Bat experiences. Thus, although all the physiological and functional facts may be used by the scientists to give rise to the feel of pain, they would not be sufficient for objectively describing/explaining what it is like to be in pain, not sufficient for rendering a nonpain-feeler with an understanding of that to which "pain" refers. The only way to come to know the latter is via the experience of feeling pain. So, it would seem that there is knowledge gained by Bat via introspection which is distinct from any and all physiological or functional knowledge about Bat to which physical science gives us access. Namely, the knowledge of what it is like for him to feel pain; knowledge (by acquaintance) which Bat acquires via introspection.

Since, if materialism were true, all knowledge obtainable about Bat would be obtainable by physical science (because identical with physiologically or functionally obtained knowledge), materialism would be false. Note too that Bat has not simply coined a private name for some aspect of his experience (although I'm not convinced that this is impossible (see Castaneda's "Consciousness and Behavior: Their Basic Connections") since the word "pain" has a causal history back to the days when others felt pain. Nor
need any of them have coined a purely private name (see ch. 5). Still, however, Bat's reference to felt pain picks out an item (cf. "an appearance") only he can understand what it is like to experience—or better, only an experiencer of pain can understand. While the scientist can say when Bat is in pain, he cannot identify the referent (as can Bat) and thus understand an essential aspect of the concept of pain, an aspect only Bat can fully understand, since nonostensive terms cannot fulfill the same descriptive role as ostensive terms.

One may be tempted to denigrate this view because Bat's claim to have this qualitatively ineffable experience appears to bear a strong resemblance to the religious mystic's claim to ineffable experience. The cases are distinct; the religious mystic's mistake (I assume) comes when he "reports" (i.e., interprets) his experience as being an experience of God (or whatever religious significance he takes it to have). Yet Bat is simply giving a name to his experience, a name others before him have given to like experiences—Bat is not interpreting (in any way near the same sense as does the religious mystic) his experience as having significance. Indeed, Bat, as the only experiencer of pain, is in much the same position as is the religious mystic relative to those of us who are non-mystics (those who have not had the mystical experience). For we non-mystics find the mystic's
attempts to describe his experience either unintelligible or (somewhat) similar to like experiences (the "runner's high") we have had. We are, I take it, willing to grant the existence of such experiences however much we disagree with the veracity of their interpretation (by the religious mystic). How much more plausible (as we see it) for the non-pain feelers (in Bat's world) to grant that Bat does have an experience that is ineffable. Bat, remember, is not claiming to experience God; he is not interpreting his experience (other than as being like that his genetic kin would have reported), but merely reporting it (he could, I suppose, even label it "Fred").

Since the actual world of which we are a part is such that most of us do have experiences (due to evolution) commonly labeled "pains," we (may) tend to think such experiences not ineffable but describable in public language. Compare (per impossible) the task of communicating such experiences to a race of non-pain feeling Martians who know all the physiological/functional facts there are to know about us. They would rightly wonder what we are trying to describe. Compare the non-mystic who follows the teachings of the mystic and comes to have the (ineffable) mystical experience; knowing how to come to have such experiences is not the same as knowing what such experiences are like--these experiences remain ineffable except to those initiates.
who can "hook-up" certain descriptions, certain bits of language to their experience.

Now the question seems to be whether or not Bat is like the drunk who "sees" pink rats: deluded. But the claim is that Bat is aware of pains, that pains are part of the content of consciousness. So, the question regarding the drunk is whether or not he is aware of pink rats. Well, are we (in the actual world) deluded when we feel pain? Are pain feelers like drunks? If so, we had better consider "saving" the drunk's appearances. For there being no pink rats in physical reality does not entail that there are no (epi-phenomenal) pink rats. Are there pains in physical reality? I don't know, but it seems that the what it is like features of pain are not, else the non-pain feeling (almost) all knowing scientist (Martian) would know what it is like to experience pain. But he does not.

To the question, "But what is left to know?", my response is that it is not a fact which can be said in English. Nonetheless, there is something left to know, as is brought out in the following. Suppose Bat is himself an all-knowing neurophysiologist who was able to rearrange the neurological structures of his contemporaries that they may come to experience pain. When properly "rewired," these individuals (we can imagine) exclaim, "Oh, now I understand what "pain" refers to!" They would then come to know what
was left out in their earlier accounts of what was happening to Bat. (A similar move would explain Nagel's point in "What it is Like to be a Bat." ⁹

Unless this pro-qualia intuition can be diffused, the plausibility of the claim that mental items are identical to (nothing but) physical items is seriously diminished—no longer would it be plausible to hold that all particular mental items are actually physical items. For without squelching this intuition, it remains plausible (though mysterious) to claim that even if pain-awareness is a physical state (say, sensing-painfully), it would not follow that the trait which distinguishes pain-awareness (from _____-awareness) is itself a physical trait. What pain **feels like** is the trait we (each in his own case) have immediate access to and by which we identify (pick out) our own (psychological) condition as the condition of being in pain. Such traits would be second-order (nonphysical, nonfunctional) traits of psychological states and we would have, in effect, dualism with respect to property instances. If one wishes to claim that there simply are no particular phenomenal items, one must at least explain why there **seem** to be. Let us consider this objection in a bit more detail.

**Evolution and Causation**

Current ontological worries have tended to drift from the conundrum of a dualism of substances to a concern
centered on a dualism of properties (that is, property instances), for phenomenal properties are arguably nonphysical in the sense that the "what it is like" feature is not captured in the explanatory net of the materialist. Cartesian related wonderment thus reappears as we focus on the nature of qualia--a robust materialistic theory of mind must not merely identify particular mental states with particular physical states (token-identity), but also explain (or explain away) these nonphysical phenomenal property instances which are part of the content of consciousness. Given that instances of phenomenal properties are distinct from instances of physical properties in the sense that "pain" is an ostensive term whose descriptive/reportive role cannot be filled by the nonostensive terms having to do with physiological or functional facts, we must ask what causal role, if any, such properties play with regard to human behavior. It certainly seems as though pain causes us to flee. With the development of modern biology, physiology and psychology, the temptation is great to invoke evolutionary considerations to help in the explanation of psychological activity. After all, evolutionary views operate within a context governed (though not driven) by physical and chemical principles, thus holding out hope for a naturalistic account of human experience. Hence, it is plausible to view our pain feeling capacities as a biological advantage in line with the theory of natural
selection. Yet all is not well here, for when pain follows an injury, is it the mental properties or the neurophysiological (physical) properties that cause the resulting avoidance (etc.) behavior that is advantageous to the species?

Not only do our commonsense folk psychological views tend to support the claim that pains, say, cause certain sorts of behavior, but evolutionary views appear to demand it. If felt pains are to be in line with the theory of natural selection, it would seem their survival value must be tied to causal usefulness. Consider the plausibility of this last claim. Given that natural selection is a relation between individual organisms and their environments which results in differential perpetuation of genes and genotypes, if felt pains were selected for, it must have been because they contributed causally to the survival of the species. This fits with our commonsense belief that pain provokes, for example, aversive reaction thus preventing tissue damage. It also appears, given commonsense and common experience (i.e., assume for the moment such beliefs are true) that the most plausible answer to the following question is a negative one: could (physical possibility) one take protective tissue-saving action upon touching (and not seeing that one has touched) a red-hot iron without having pain sensations?
Some have taken such ruminations to indicate that interactionism makes felt pain an understandable biological necessity (e.g., Puccetti in "Is Pain Necessary?"10). Understandable because on this view felt pains have a reason for being, namely, they are causes of certain behavior advantageous to the species. But such interaction is itself not (plausibly) understandable for it violates scientific conservation principles; not, however, because of a loss of energy when physical items cause mental items, for physical energy need not be required to bring about a mental occurrence (See Cornman, Lehrer, Pappas, Philosophical Problems and Arguments.11) Rather, the problem lies in the notion of the causation of the physical by the mental, for in this case it does seem that there is a violation of scientific principles. Consider this problem as viewed by Michael Levin:

Another traditional argument, however does show decisively that the causal relations between the mind as the dualist conceives it and the body are impossible. Originally my leg, which has mass w, is motionless. Then it moves a distance d. Enough work was done to move a mass w a distance d. Where did the energy come from? It is an empirical fact, the law of the conservation of energy, that the energy had to come from somewhere, and not only does it come from some preceding physical event, no immaterial substance could possibly supply mechanical energy. The mind can apply no physical force to the leg, while the contraction of the hamstring muscle
supplies just the right amount—and the bodily event just is the motion of an object of mass \(w\) through distance \(d\). Thus the contraction of the muscle is the cause of \(e\). If there were non-physical causes of physical events, the energy in the universe would increase.¹²

At best we are left with a case of inexplicable causation. One can, of course, "dig in" and simply admit that such causation is mysterious and conclude, as does Puccetti, that "Interactionism is the worse possible mind-body theory, except all the others."¹³ This conclusion can be avoided, however, if another theory can account for our commonsense beliefs without running afoul of accepted scientific principles (as well as the phenomenological data). Epiphenomenalism does appear to account for our beliefs that pains do (typically) provoke aversive reactions and that typically one cannot take aversive action without having a pain sensation. This theory is not, however, free of mystery.

Suppose that mental property instance \(+P+\)—the particular "what it is like" feature of pain at \(t\)—was an ontological mule, a mere epiphenomenon. Let us represent this state of affairs as follows:

\[
\begin{array}{c}
I \\
N_1 \quad N_2 \quad N_3 \\
\end{array}
\quad \quad +P+ \quad A
\]
Here the solid line represents the "causal path" leading from injury I (of an organism) to avoidance behavior A. \( N_1 \) will then represent the neural complex that is the proximate cause of the pain quale \(+P+, \) a purely mental item. \( N_2 \) and \( N_3 \) are further neural complexes which, as causal factors, maintain the causal link between I and A. Now, on the assumption that the capacity to be in states such that one feels \(+P+, \) that of which we are aware in experience, evolved via the principle of natural selection, it would follow (it seems) that \(+P+, \) in order to comport with commonsense, must contribute causally to the production of A and hence, to the survival of the species. But the above representation makes it clear that the cause of A rests solely with the neurophysiological components. That is, there appears to be no reason why blocking the production of \(+P+, \) would stop A from occurring as long as the causal path through the neurophysiological factors remained intact. (Of course, if \( N_1 \) is physically necessary for the occurrence of \(+P+, \) and vice versa, then you can't have one without the other; see below.) Nonetheless, initially at least, the data of introspection makes more plausible the homespun belief that it is the hurtfulness, the "what it is like" feature of pain that is responsible for my avoidance behavior. For when asked what caused me to withdraw my hand and apply salve, my introspectively based answer is that it hurt—the felt pain caused me to do as I did.
So, in a clash between the epiphenomenal status of pain and an evolutionary account of the origin of pain, the victor is often thought to be the latter. J. L. Mackie, however, proposes that mental features are automatically selected along with their neurological basis and that the two are causally inseparable. In effect, then, it is the neurophysiological base that is actually selected (due to its causal usefulness). Such causal inseparability, it is alleged, accounts for our belief that the mental causes, or causally effects the physical—the belief that the hurt causes the avoidance behavior. Yet this belief is, in an important sense, false, for only the neurological items are causally efficacious. The mental is epiphenomenal and not causally efficacious although, if there are neural correlates of felt pain which are causally responsible for particular avoidance behavior at \( t \), there will be felt pain at \( t \). Thus, on this view, if there are no felt pains at \( t \), there are no neural items responsible for avoidance behavior at \( t \) (and hence, no avoidance behavior at \( t \)) and if there are no neural items responsible for avoidance behavior, there will be no felt pain—that is just the way Mother Nature, as a matter of fact, operates. (As we will see below, this claim seems to amount to supervenience or emergence of the mental upon the physical.)
The logic of this claim cannot be faulted, though it is not clear why there is this nomologically necessary connection between felt pain and certain correlative neural items. Since such a view appears to make felt pains utterly gratuitous to our evolution (logically speaking we could have the avoidance behavior, etc., without the hurt), we may simply be trading mysteries with the interactionist. If epiphenomenalism is true, then for some unknown reason we experience pain which is not causally efficacious in bringing about avoidance behavior although, since it is causally inseparable from that which is causally efficacious, we believe it to be causally efficacious. My own inclination is to side with the epiphenomenalist's mystery for it does comport with common experience (the phenomenological data) and does not violate accepted scientific principles--it is just a brute fact that Mother Nature could not (physically), as Lycan has put it, have built a successful damage-avoidance system like ours without in the process creating the capacity for felt pain. (This notion of the "capacity for felt pain," however, sits ill with me--I do not take felt pain to be a disposition.)

We are, then, faced with a quandary; since there is some reason to think phenomenal property instances distinct from physical property instances, where the latter appear to be sufficient causes of bodily movements, it is far from clear just what causal role the former could play. If they play
none, we are left with the presence of an ontologically unexplained residue. On the other hand, to attribute a causal role to qualia as the property instances of states which are responsible for the states' causal powers is to flirt with interactionism and its well-known difficulties.

It is this general problem that the functionalist (as materialist) must overcome. The functionalist attempts to combine the identity theorist's insight— all actual mental particulars are physical, thus providing a causal account of interaction while avoiding epiphenomenalism—with the behaviorist's insight— a relational interpretation of mental properties is called for— by arguing that a mental particular is of a certain type (e.g., a searing pain) if, and only if that particular plays a specified role in the mental life of an organism (system). The materialist thrust of this position lies in the empirical claim that only physical items will realize (play the role of) the functional properties that define (or are identified with) mental states (types). What these functional properties (construed as types) are not, and hence, since functional properties = phenomenal properties (types), what phenomenal properties are not, are intrinsic nonrelational items. What they are are relational behavior-causal properties explicated as functional syndromes relating sensory inputs, internal states and motor outputs. Thus the functionalist claims that the
concept of a causal role is to be explicated in terms of the causal relations one mental state has to other mental states—a pain, say (construed as a type), is to be individuated with respect to its causal role. Though functionalism, like behaviorism, is relational, it is important to emphasize a crucial distinction: the latter is concerned only with the input-output syndrome (thus the "black box") whereas the former also stresses inner causal processes—as far in as our discerning abilities will take us. Given that understanding of functionalism, however, a glaring problem remains: the particular pain I feel at t is obviously a functional property (instance); yet if the particular pain (by virtue of playing the appropriate functional role) is a neurophysiological (physical) item, how is the phenomenological aspect of my pain (which I argue is essential) explained or explained away? Indeed, how is this problem different from that facing the identity theorist? Let us investigate.

**Emergence/Supervenience**

So, to be in pain an organism (system) must have a (physical) structure so organized that it can be said to realize a functional profile identified as the pain profile by its functional properties. Where such a realization occurs, a particular individual pain quale may be said to
emerge or to supervene from the complex organizational structure of which the functional characterization is true. On the emergentist view, we may think of the pain quale as an indefinable "holistic" attribute of the organism, which is correlated with a reducible attribute definable via the physical properties and relations of individual systemic items.\textsuperscript{16} For the emergentist, then, the pain quale is not reducible to physical properties of_____ but is (merely correlated with) the basic-level individuals of physics. Supervenient views treat the pain quale as an independent object sharing the same region of space also occupied by the physical items realizing the functional syndrome and nomologically correlated with those items (we will return to this in ch. 4). Now, I won't pretend to fully understand these two views nor attempt to point out how they differ. Rather, I shall simply point out that they appear to be of the same logical type as is the epiphenomenalist view. To this end, let us delve deeper.

Either of these last two views might be thought preferable to the epiphenomenalist claim that the pain quale is a non-physical particular. Both, however are subject, claims J. Rosenberg\textsuperscript{17} to a similar criticism: neither allows for causal explanation since they do not allow that qualia play an essential role.\textsuperscript{18} As we have seen, however, such an objection is hardly decisive; the real difficulty is
ontological in that we appear to have a non-material residue left over even after the causality is explained. As Keith Campbell puts it:

Emergence theory maintains that the appearance of non-physical properties in neural systems is not a cooperative effect of properties of their simpler parts. It is a quite new phenomenon, which just emerges in a certain sort of physical system. There is not, and cannot be, any explanation for such a fact. If it is a fact, it is a brute fact.19

Let us pursue this claim and strengthen its apparent plausibility. Considering, for the sake of simplicity, the structure upon which pain is said to supervene a neuroanatomical structure, we can represent the claim thus (for the present point the alleged difference between emergence and supervenience can be ignored):

\[ I \rightarrow \begin{array}{c}
+P+ \\
N_1 \\
N_2 \\
A
\end{array} \]

Again, the solid line represents the "causal path" from injury I to, in this case, neurophysiological complex N₂. Supervening on N₁ will be the pain quale, +P+, while supervening on N₂ will be the perceived bodily movement that constitutes avoidance behavior A. It is to be noted that the claim that +P+ is relational only seems relevant when +P+ is construed as a type. For as a type, +P+ is picked out,
by one not experiencing the pain, via its relations to inputs, internal states and outputs. But present concern is with the particular occurrence of pain I feel now—I do not identify my pain in any such way. Thus, although another may identify my pain relationally, I do not. Now, if in this particular case the supervenience relation is one of biconditionality, it may be thought that any justification for claiming $N_1$ to be the cause of $N_2$ will also serve

mutatis mutandis to justify attributing a causal relation between $+P+$ and $A$. But, as we have seen, epiphenomenalism can be understood in this very way and that our pro-qualia intuition concerning the belief that $+P+$ is causally efficacious can be explained while still leaving an unexplained residue. Thus I, for one, do not understand this short of claiming that supervenience is reduction. Consider a case which flushes out an important feature regarding our intuitions about causality.\textsuperscript{20}

Imagine watching a game of billiards and observing the lawlike regularity of a well played game. Certainly the law-like regularity, the constant conjunction between the (apparent) impact and consequent movement of balls justifies us, ceteris paribus, in claiming that the motion of one ball causes the consequent motion of another. Surely we have here a paradigm cause of causality. Yet our intuitions jar us into denying cause-effect when the ceteris paribus clause
is violated as follows. Suppose the billiard table so rigged that once a ball is in motion it falls under the control of an electronic magnet wielder. Never is there any contact between balls; they all stop at an infinitesimally close distance from one another with the resulting motion being due to magnetic factors. Our intuitions, I take it, lead us, in the case of the rigged table, to deny that one ball causes movement in another—despite the law-like regularity and observed constant conjunction. Though the latter characteristics may serve to warrant the attribution of a causal connection, they are not constitutive of that connection. Rather, unless something (e.g., motion, energy) is imparted from one ball (object) to another, we withhold our attribution of causation (at least if we know or believe otherwise). Let us apply this intuition concerning imminent causation to the above diagram.

As in our earlier diagram, it seems clear that a causal story can be given that will take us from $N_1$ through $N_2$ to $A$. Mere correlation (no reduction) of $+P+\rightarrow A$ with $N_1-N_2$ will not warrant us in attributing a causal role to $+P+$ unless we have reason to think something imparted from $+P+$ to $A$. But since, lacking reduction, the pain quale is not itself a physical item, it is only supervenient upon physical items, it is a mystery how it could impart anything to $A$ via $N_2$. Furthermore, while the supervenience relation
between \( N_2 \) and \( A \) seems to be a constitutive relation in the sense that a causal story can be told which would trace the bodily movement as the outcome of neurophysiological happenings, the same is not obviously true of \(+P+\). Unless one wishes to identify \(+P+\) with \( N_1 \), it would not make sense to claim that the supervenience relations between \(+P+\) and \( N_1 \) is constitutive. \(+P+\), again, appears as an ontological mule, for it does not cause (serve as causal factor for) \( A \).

In short, a causal explanation for the occurrence of \( A \) need make no reference to \(+P+\) and the supervenient picture appears epiphenomenal in form. Notice that construing a particular instance of the pain quale as a functional property makes no sense, for the former is a token, the latter a type. The main question centers on the specification of realizers of functional properties: what are the property instances that enter into causal relations? Again, our concern is with particulars (tokens), not universals (types). Thus, the functionalist-qua-materialist must provide qualia with a causal role as robust as that understood to obtain between neurophysiological (physical) items if he is to be other than an epiphenomenalist and if he is also to account for the phenomenological data. Let me, then, set the problem facing the reductionlistic-minded philosopher from the point of view of emergence, the nasty emergence alluded to by Campbell. Rather than being overly schematic, I will be
concerned with the pains that we as humans often feel.

First, let me attempt to make clear the sense of "emergence" that I find troubling, for this is not an easy concept with which I deal. Of course there is a sense in which emergence need not be taken as nasty. If we start with a pile of bricks and organize them in the appropriate way, according to an organizational plan or blueprint, a building emerges. This is hardly a cause of worry for we can see how the building is a cooperative effect of properties of its simpler parts, even though, when reduced to mere bricks, the building as such is lost. (But of course we know all along that the building just is the bricks so arranged—the latter are constitutive of the former.) This, however, is not an apt analogy for emergence with regard to the nervous system and the appearance of non-physical property instances like the particular pain I feel now. Particular instances of felt pain may arise from a certain organization of the neural system, but it is not at all clear how these pains, as elements of being, are to be likened to emerging buildings.

When we label the organized collection of bricks a "building," we are thinking of nothing but these bricks—they are constitutive of the building. But the pain I feel now is not plausibly thought of as nothing but certain neural items appropriately organized; there is more, namely, how it feels. The disanalogy is obvious; in the case of the
building all we have are bricks. When felt pain emerges from structure we have a something which is entirely new and, unlike the building, the structure upon which it emerges is not constitutive of the felt pain. The arguments which follow make this clear and hence, help make the sense of "emergence" which troubles me clear.

If pain is a property of humans, then pain is reducible if each instance of pain is itself reducible to instances of micro-properties of the constituents of humans. Should the connection between instances of pain and instances of micro-properties of the constituents of humans be no stronger than mere correlation (as appears to be the case on the epiphenomenalist view), then particular instances of pain are emergent. (Such micro-properties, of course, are to be construed as purely physical.)

We may wish to think that pain (construed as a type) is instanced in various ways, e.g., shooting/throbbing/searing pain at \( t \). It seems clear that these phenomenal property instances are not physical property instances of physical/functional properties (types). (Arguing for this is, of course, the task that I have set for myself.) Yet for materialism to be true, for epiphenomenalism to be avoided, these former property instances must be nothing but the latter property instances. So even if one wishes to make the type-identity claim (see below) that pains = functional items,
it will not automatically follow that pains are nothing but functional items since phenomenal property instances will be emergent, not being reduced to physical property instances.

Part of the reason for maintaining this emergentist thesis is that particular instances of pain cannot adequately be described in physical or functional terms, for "my pain at t" is an ostensive term which refers to (which I use in order to make identifying reference to) that aspect of my experience to which only I have direct access. Thus only an experiencer of felt pain can understand that to which "pain" refers, only one who shares my "form of life" can understand such language. True, another must rely on behavioral clues (overt or covert) in order to truly predicate "pain" of me. Foremost for another's truthful predication of "pain" to me is my sincere report that I am experiencing pain. It is this view, I shall argue, which is the most plausible view to adopt regarding qualia. Unfortunately for the materialist, this view cannot be made consistent with materialism. "Cannot be made consistent" in the sense that it is conceptually necessary \(^{22}\) that qualia be so understood.

One may wish to talk of "conceptual revolutions" and the like in order to avoid this result; I shall not argue in detail against such a move since I fail to see any plausibility in such claims. Indeed, I tend to view such a move as an admission of defeat being the final parting shot of the
retreating materialist—it reminds me of the skirmish between Moore and the sceptics: I know when I am in pain and what pain is like. And it is the "what it is like" feature which escapes the materialist's net. By way of methodological note, I must make it clear that I choose to talk, not about language, but about features of the world. In a sense, then, I may be committed to "the given" in the following version:

P's are given =df P's are the entities [property instances] which are experienced even before learning and independent of language, and with which a language must be associated if it is to be learned.23

That this move is warranted will, I hope, become clear from the arguments that follow. At any rate, I am close to maintaining it as a postulate. I must confess that a good deal of "Sellarsian materialism" is not discussed herein. Some of my reasons will, I take, become clear—I move from a different starting point. More to the point, however, I am not doing a study on Sellars, a task for more than one thesis.

As will become evident in Chapter 4, supervenience is to be construed as reduction, where claiming that a psychological (mental) description supervenes on a physical (neuroanatomical) description amounts to claiming that psychological language is simply a way of interpreting at a molar level the significance of behavior and the significance of
inner happenings in terms of the contributions they make to behavior. We must, of course, bear in mind the ontological decisions that must be made for this to be plausible. After all, to interpret a functional state as, say, a pain state (type) requires that we answer two questions: (i) what sort of functioning is appropriate to (felt) pain?, and (ii) why is it appropriate to (felt) pain? To answer (i) and (ii), we must make reference to behavior, including verbal reports. What such reports make reference to is not obviously physical. Supervenience as reduction, however, does not appear straightforward (going from mental descriptions to physical descriptions in one fell swoop), but rather involves topic-neutral description (see below). This "pain" is reduced by employing a topic-neutral description: a pain is whatever state typically produces (causes) pain-behavior. This makes for a causal account of pain in which physical items, it is contended, will be discovered to be the items which typically produce pain-behavior and in which introspective awareness will involve awareness of only physical items. Such topic-neutral "reduction" opens up the possibility that possession of physical properties is sufficient for the truth of mental (construed as topic-neutral) ascriptions. To be successful, however, it must be that there is nothing grasped non-neutrally in experience; it is a major contention of ensuing chapters that topic-neutral analyses of felt pain will not suffice.
Token Identity: Types and Tokens

To fill a lacuna, we need to consider another representation where a particular instance of felt pain, +P+, is assumed identical to neurophysiological items. As Jay Rosenberg states with regard to this reductive materialism:

it subscribes to the thesis that the only objects having ultimate ontological status are "atoms in the void", and thus rejects both substantial "sensoria" and epiphenomenal "sensa". But reductive materialism attempts to go further in proposing that the only actual items or entities having ultimate ontological status are these "atoms in the void", and thus rejects as well the idea that there can be states of a person which are not complex motions of such atoms (although, of course, "lawfully correlated" with such motions). 24

Consider the following diagram:

I ___________________________________ A

N_1

N_2+P+

What are we to make of calling the observed biconditional relation between N_2 and +P+ (contingent or a posteriori) identity? One response would be to interpret the observed biconditionality as strict identity (i.e., reductive, nothing but identification) where N_2+P+ is purely physical (N_2) although sometimes talked about under another
guise (+P+). Then there would be no particular instances of felt pain distinct from neural (physical) items (only the appearance of such?). But this eliminativist view, I argue, is implausible: there are particular felt pains which are part of the content of consciousness (that we are directly aware of) and neural items are not part of the content of consciousness unless that of which we are directly aware is a property (instance) of such items. Such property instances, the appearances, must themselves be explained. (I explain and argue for this below.) Moreover, merely changing the labels we attach to these contents does not eliminate their occurrence—felt searing pain by any other name is still felt searing pain and "searing pain" or its replacement can only be "defined" ostensively. Thus, saying pain (universal) supervenes on certain levels of organization (we can attribute pain if enough complexity, etc.), and that particular felt pains are identical with the realizers of certain functional syndromes, runs into serious difficulties if the pro-qualia intuition cannot be diffused. Justifying the claim that qualia really do present these problems is the long task ahead. To begin, let us ask what is essential to, say, pain.

Explanation, lastly speaking, is to aid the understanding by providing answers to certain questions. In the present context, then, consider: what is the essential as
contrasted with the accidental element of being in the state of pain? What, that is, is the property or feature necessary to the nature of pain? As we have seen, contemporary functionalist theories maintain it is a relational property characterized in terms of sensory inputs, internal states (abstractly considered), and motor outputs: (roughly) to be in pain is to be in a state which plays a particular causal role in a larger functional story. Another way of putting this question is by considering what two organisms have in common when both are in pain. What determines type identity qua psychological state? To see the relevance of this question, consider a modification of Jerry Fodor's distinction (in *The Language of Thought*) between type physicalism and token physicalism:

**Type Ø**: every property mentioned in the laws of any science is a physical or topic-neutral property.

**Token Ø**: all the individuals that the various sciences talk about are physical individuals.

Now, in seeking the means by which we identify a psychological happening as being of a certain type, we are asking in virtue of what are certain psychological happenings all of the same kind. William Lycan (see ch. 4) in pursuing a reductive type-identification of mental items with functional states, espouses Type Ø. For two organisms, \(O_1\) and \(O_2\), to both be
in pain is for a functional description to be true of both $O_1$ and $O_2$; since this functional description is topic-neutral, there is a reduction from mental type to functional type. Token $\emptyset$ enters with the claim that the items playing the causal role in the functional syndrome will all be physical items. Hence, the importance of the claim that there be no phenomenal (non-physical) property instances of which we are aware. A caveat is in order here.

In a sense posing the question as above may raise a bit of a red herring. If the feature in virtue of which a given psychological state is of a kind, say, the pain kind, is considered an essential property of the state, then should we agree that the feel associated with pain is an accidental feature, not essential, it hardly seems that functionalist woes are thereby overcome. While the admission that the feel is not an essential property of the pain state would cast doubt on the epistemically privileged position typically granted pain reports—Morgan may be in pain even though (since he doesn't feel pain) he reports he is not—it may well be an admission the pro-qualia defender can ward off as not damaging to his position. Concern, after all, centers on pain when felt. Thus, even if the feel were an accidental (not essential) feature of the pain state, it would be, when felt, an item in need of explanation. Since particular pains are felt, that which is felt is in need of
explanation. So, when Morgan reports the experience of feeling a searing pain, what is he referring to?

Suppose we allow our concept of the pain state to be bifurcated thus: state A: the state of being in pain simpliciter; state B: the state of being in pain and feeling it. The functionalist, of course, claims to have an adequate explanation of the difference between states A and B: the functional syndrome will differ in each case. Yet the pro-qualia intuition has it that there is more than just a functional difference: what it is like to be in state A is significantly different from what it is like to be in state B. One can know what the latter is like only via experience (realization of the state B functional syndrome) and can only learn the referent of the locution "felt pain" ostensively (with, we might say, a little help from our friends—members of our language community, see ch. 5). Indeed, that which is ostended is the "what it is like" qualitative feature arising from the complex inner workings which constitute the realization of the relevant syndrome. But, where it is this difference in realization of different functional syndromes which gives rise to the felt difference between A and B, we are back to emergence or supervenience. So, though I am willing to agree that a person (system) can be in pain (state A) or feel pain (state B) only if they realize the appropriate functional profile, I am not willing to accept
the view that all there is to being in pain is explained once the functional syndrome has (or the realizers have) been specified. Should one wish to argue that functional explanation is all the explanation that can be given, then in lieu of an adequate causal/ontological account of materialism will be difficult to reckon with the facts—such explanation would not be explanation enough. With this caveat entered, let us return to the question of what it is in virtue of which a psychological happening is of a kind—remembering that, in the case of pain, the relevant pain state will be of type B.

Given an explication of the concepts of "law" and "theory," Fodor maintains we may roughly characterize the kind predicates of a science as those whose terms are the bound variables in its (proper) laws. Fodor, in arguing that classical (positivistic) reductionism is, for empirical reasons, too strong a demand for the unity of science and that Token Ø will do, denies that type-to-type correlations between psychology and neurology are needed. As materialists all we need do is pair every psychological individual with some neurological individual or other. We do not need to pair types of psychological states with types of neural states as long as the token-realizations of the psychological states (such states being picked out according to their functional roles) are physical. For the
functionalist, then, the generalizations of psychology are most naturally couched in terms of a level of abstraction in which functional role determines type identification, not physiological composition. Thus the functionalist is able to account for, e.g., (neurophysiological) equipotentiality (as well as avoid the charge of species chauvinism); neural item \( N_1 \) in Jones at \( t \) can be identified as pain in Jones at \( t \) even though neural item \( N_2 \) may be identified as pain in Jones at \( t_1 \) and \( N_1 \) identified as fear. (This is, of course, simplistic but the idea is clear--functionalism need not run counter to such empirical discoveries.) Being in pain (type) is not identified as \( N_1 \) (or \( N_2 \)) even though any particular occurrence of pain (token) is some physical item or other (or, so argues the functionalist as materialist). It begins to look, then, as if reduction must be of particulars (property instances); \(+P+\) (if not merely eliminated) is to be reduced to (reductively identified with) some physical item or other.

Thus, asking what determines type identity \textit{qua} psychological state, is asking what properties are relevant for an \( n \)-tuple of psychological happenings to be of the same kind. To embrace materialism, I have argued, these properties (construed as individual property instances) must all be physical if we are to avoid a variety of psychophysical dualism of properties (in particular, it seems, epiphenomenalism as discussed above). A response often heard is that
this type of dualism is no more of a problem for the materialist than is the "dualism" to be found in auto engines. In an auto engine it seems quite obvious that the "sparks" of ignition are the causal result of the purely physical items constitutive of the engine. If it is then argued that there is no dualism in the case of the auto engine, no mysterious élan vital, we are apt to quickly agree. But we must be careful what we are agreeing to; it is true that there is no dualism in the auto engine even though the "sparks" may appear to be non-physical property instances emergent upon the physical constituents. But, of course, the "sparks" do not appear as such to the engine (?), but to us. Indeed, such "sparks" are the way the physical workings of the engine appear to us. It is the relation between these appearances and, say, human (conscious) observers that is the locus of the problem just as it is the relation between felt pains (after-images) and the physical constituents of the central nervous system—how those constituents appear to us—that sets up the phenomenal property (instance) difficulty. So, it is true that auto engines are not, in and of themselves, dualistic (they are not conscious) systems—it is in their relation to human (conscious) observers that dualism threatens. Given that the reduction the materialist needs requires that the items (property instances) of which we are aware be physical, it
is far from obvious how the phenomenological data is accounted for—what are we to make of the way in which physical activity appears to us, what are we to make of the "sparks" and felt pains? The congenital non-pain feeler with total physical/functional knowledge will not have access to a fact known by the pain feeler. I turn next to bolstering this claim and to showing how implausible it is to think the materialist can, via his reductive strategy, account for the phenomenological data, the what it is like of felt pains.

Neutralizing Introspection

Although it is, in fact, obvious that we do experience pain, e.g., touching a red-hot iron typically hurts, materialists often maintain that this experience is nothing but overt behavior (naive behaviorism) or covert behavior (reductive materialism). Speaking epistemically, their attack is most often directed toward the source of the belief that the felt hurtfulness is other than behavior (overt or covert): they attack introspection as a legitimate source of reliable information (knowledge) regarding the intrinsic nature of, say, the pain experience. Our claims, first-person reports, about such experiences, e.g., "I am aware of a searing pain in my hand right now," it has been maintained, are actually (really!) reports about (I bypass naive behaviorism) covert behavior,^{20} or inner brain (physical) activity.^{29}
Thus, it is alleged, first-person reports do not attribute phenomenal (non-physical) qualities to our experiences; that is, what we are aware of in introspection must be purely physical. To make this plausible the materialist must claim that first-person reports attribute to the physical items constitutive of felt pain only physical qualities or else such reports do not attribute any particular kind of quality at all.

Such a view has interesting consequences and makes for uncomfortable questions. It would seem that the only covert activity essential to felt pain is that activity of the CNS (the core person) which is a necessary and sufficient causal condition for felt pain. Now, then, how is it that we pick out the relevant behavior in the CNS that is causally necessary and sufficient for felt pain? That is, what is a brain item's being experienced over and above its occurring? How is it that the what it is like feature—the hurt—is to be recognized as present by the materialist? It may be that we recognize the felt pain of others (come to know they are in pain) by noting their behavior, and given present and future correlations with their reports we may do this by noting inner activity of such and such a sort. Surely, however, this is not how they recognize that they are in pain (nor how we recognize that we are in pain) nor, I argue (see ch. 6), could we truly recognize that they are in pain if they
could not so recognize—correlations depend on this. We must ask: what makes "pain now" predicable?

Smart's Tack in "Sensations and Brain Processes," was to maintain that first-person reports (descriptions) were topic-neutral. For a statement to be topic-neutral is for it to pick out some item without ascribing any intrinsic features to the item picked out. For an item to be known topic-neutrally would be for it to be known under a topic-neutral description (and, it would seem, no other non-neutral [intrinsic] description). Such descriptions need not be totally neutral, but neutral relative to a context (e.g., "whoever is the murderer of Smith," "whatever is responsible for the grapes rotting), where we have a sense of the context. In the case of felt pain, however, this is a radical move for the context surrounding such topic-neutral descriptions (e.g., "something is going on in me which is like [public event]") does not appear to invite identification with the physical—which it must if topic-neutral descriptions are, as is alleged, stopgaps for (future) physical descriptions. After all, which internal states count as experiences of pain?—those of which we are aware. Do we gain no non-neutral knowledge about the nature of our experiences from such introspective awareness? The materialist must claim we do not. So the topic-neutral account was presented to do justice to the claim that first-person reports are reports: topic-neutral reports. The
object of experience, the what it is like feature, is topic-neutral reported and such experience only gives us topic-neutral knowledge of the objects of experience.

The functionalist, in picking up on this notion of topic-neutrality, is wise not to claim that functional (topic-neutral) descriptions and mental descriptions are synonymous. Nor, is it alleged, need he, for the relation between the two sorts of description need not be synonymy but only that of reference-fixing. Thus, it is claimed, the reference of mental expressions are accidentally fixed by topic-neutral descriptions and "pain" rigidly designates, e.g., whatever it is that happens when [public event]. Future science will tell us what that is. On the functionalist account, my pain at t = whatever physiological (physical) state (future science tells us) I am in at t that is of the functionally specified kind. Mental language, construed as itself being topic-neutral, amounts to but heuristic reference-fixing description.

On this view, my immediate acquaintance with (felt) pain does not guarantee that I know any descriptions distinct from those others know. And all such descriptions are topic-neutral couched in terms of public events. Only our ignorance of the inner workings of the core person prevents us from describing, in explicitly physical language, that which we fix the reference of with our mental descriptions (first-person reports). Therefore, what makes "I am in pain at t" true is an open question for future science to answer. The question, "what do I know about my pain?", is answered by the functionalist with, "very little and nothing essential." Thus our topic-neutral descriptions (first-person
reports) are mere stopgaps covering our ignorance about the true nature of our inner states. Since, I take it, this is (at least) *prima facie* implausible, why think it an adequate account of the phenomenological data? By the end of this study I hope to show it not plausible at all.

It may be thought plausible because it may be thought that there is no fact about (felt) pain in $S$ at $t$ known by $S$ but not known (knowable) by others, because facts must be capable of linguistic expression. Thus the person incapable of feeling pain but who knows all physical/functional facts knows all there is to know about (felt) pain. This seems simply false. True, felt pain cannot be described other than neutrally (and couched in terms of public events). But those of us acquainted with (felt) pain understand these mental descriptions because we know what it is like to experience pain. That is, when the reference of "pain" is fixed by some topic-neutral description we are not in the dark, we do understand what is essential to the experience of pain even though we may not (indeed, do not) understand the total causal story of how felt pain is produced. If one wants to maintain that there is no fact about (felt) pain distinct from all physical/functional facts because such facts must be capable of linguistic expression and all descriptions of (felt) pain are only compatible with physical/functional, i.e., public states of affair, then so much the
worse for that view of facts. Moreover, this appears to but draw a pink skunk across our track; there is something (call it what you may) that a pain feeler knows about felt pain and which enables him to understand what another is going through when that other sincerely reports he is in pain. Someone who has never experienced pain does not know this aspect of pain and does not understand what another goes through when he truly reports pain. To claim that what is known by the former (pain feeler) is not a fact, simply ignores the data, especially when combined with the claim that the "all knowing" non-pain feeling scientist knows all (the facts) there is to know about pain and hence, knows what is constitutive of felt pain.

Isn't it much more plausible to hold that there is a (non-physical) property instance that justifies my use of "I am in pain"? It seems implausible to maintain that I am trained to use a referring expression, e.g., "something is going on in me which is like what goes on when I touch a red-hot iron," without noticing an aspect of my inner state to which reference is made. To reply that "what is going on when . . ." refers to a brain (physical) item, that what is discriminably present is a brain item, and only a brain item (how does it appear?) simply ignores what it is like for me to be in that (to realize that) functionally defined state. I contend that what it is like for me cannot be ignored
(i.e., neutralized) on pain of ignoring a datum, namely, the feature by which I identify (pick out) that state as being a pain state of mine. (And, by way of correlation, by which another picks out my state as a pain state.) Certainly when I feel pain, I do not merely "observe" inner happenings (something I know not what) which generally precede or accompany my yelling, reaching for ointment, etc.—I know all too well (essentially) what I am aware of. Why maintain that that which I experience is topic-neutral-to-be-discovered-to-be-physical because my report (the description I use to inform you) is topic-neutral? "Pain" is defined only ostensively. To maintain otherwise rightly meets with howls of outrage.

Suppose Smith has total physical/functional knowledge of the pain feeling Jones. Smith, by way of explanation, provides us with complete physical/functional description of Jones when the latter is in pain. He explains how certain neurons, composed of certain chemicals, composed of certain molecules . . . constituted by atoms in the void behave in certain ways as to have certain (likewise fully described) effects, which eventually lead to Jones' pain behavior, including Jones' report that he is feeling pain. Smith, however, has never himself felt pain since he has an abnormal "wiring diagram." We may thank Smith for his information, which is extremely useful for prediction and
control. Yet we will point out that it is not complete explanation, although complete causal explanation, and that this physical/functional descriptions are themselves topic-neutral, mere reference-fixing heuristics serving as stopgaps to cover his (Smith's) ignorance of what (felt) pains are like. We will point out that once he manages to reconstruct his "wiring" so that he too may experience felt pain, he will discover that felt pain is more than what he has claimed, more than the causal story (complete with physical role occupants) tells us. In anticipating the future reconstruction of his CNS, we anticipate that Smith will gain new knowledge, new understanding about the intrinsic nature of felt pain; he will come to know what it is like. He will, of course, be unable to describe the new experience in other than topic-neutral-cum-public-event descriptions which we understand and which he now understands—ostensively.

Now, which story seems more plausible: that we (pain feelers) do not know the intrinsic nature of felt pain, that our pain reports sincerely uttered are mere topic-neutral reference-fixing heuristics serving as mere stopgaps and that future science (Smith) will tell us what the intrinsic nature of pain really is, or that Smith, despite total scientific knowledge of the physical constituents of humans, has only topic-neutral knowledge of pain and will, when he
comes to experience pain, discover what the intrinsic nature of pain really is?

I claimed that the topic-neutral account was presented to do justice to the claim that first-person reports are reports--topic-neutral reports of experience. The object of experience, it was claimed, is topic-neutrally reported with the result that experience only gives us topic-neutral knowledge of that object. Given the above, we can see that this account does not do justice to the phenomenon of introspective awareness of particular occurrences of felt pain. As I see it, introspective awareness enables us to gain non-neutral, non-physical, phenomenal knowledge about the essential nature of our experiences and that this view is more plausible than the view that science will tell us that the essential nature of our experiences is (in all cases) physically non-neutral--that physical knowledge is total knowledge. If it should be maintained that introspective experiences, e.g., felt pains, have two sorts of essential characteristics, functional and phenomenal, then, while introspection cannot provide us with complete understanding (who claims it does?), science cannot give us the complete story and materialism fails by being an incomplete (and not completable) account of human experience.
As with any ism there are many gaps and divides among the functionalists; my strategy in what follows is to examine how different functionalist theories attempt to answer further questions raised by the functional syndrome response to the question of type identification of psychological states. In this way it will be possible to (i) delve further into current functionalist views, thereby expanding on intra-theoretic differences, and (ii) examine some qualia based objections that pose serious problems for any current functionalist theory—indeed, for any theory of mind, since such worries are indigenous to the mind-body problem. I shall argue that the juggernaut of science has not—not yet—squelched the intuitions that generate the apparent intractableness of this problem, by showing that the general ontological/causal difficulty raised above remains the unsolved mystery it was for Descartes. In so doing, some erroneous objections to functionalism—founded on a misunderstanding of the position—will be put to rest, thereby focusing attention on the truly difficult questions facing contemporary materialism. As I proceed, issues only touched upon above will receive more detailed examination.
2. INVERTED QUALIA

According to this difficulty, modeled on the inverted spectrum puzzle, functionalism like its dame behaviorism fails, it is claimed, to grant sense to something that does make sense and thus leaves something out in its account of human experience. We have seen how the functionalist proposes to type identify psychological states; the current objection is that with regard to some psychological states this proposal fails to accommodate a necessary feature of type identification, namely, the qualitative or "what it is like" feature of such states. Consider Ned Block's posing of the objection:

The Inverted Qualia Objection can be introduced by attention to the familiar inverted spectrum hypothesis, the hypothesis that though you and I have exactly the same functional organization, the sensation that you have when you look at red things is phenomenally the same as the sensation I have when I look at green things. If this hypothesis is true, then there is a mental state of you that is functionally identical to a mental state of me, even though the two states are qualitatively or phenomenally different. So the functional characterizations of mental states fail to capture their "qualitative" aspect. . . . If inverted qualia are possible, then mental state 0 is not identical with any functional state, and if functionalism claims that every mental state is a functional state, functionalism is false.
Although functionalism does claim that every mental state is a functional state, I shall show that the mere possibility of inverted qualia does not suffice to show the theory false. Rather, the most plausible response to this objection leads to a much deeper difficulty—the ontological/causal conundrum discussed in Chapter I. It is not, then, clear that this problem does not point out a problem for functionalism as materialism. Culling from several recent accounts, let me put the challenge as follows.

Seeing Red What is Green

Suppose the nefarious Dr. Z performs prenatal surgery on the twins, Sid and Amy, replacing Sid's cornea with an exact replica except for having a color inverting lens. It is not difficult to imagine, as it turns out, that there is a psychological difference—a difference in "mental state"—between Sid and Amy when we (being unaware of Dr. Z's antics) would least expect it. When he sees green she sees red, when he sees blue she sees yellow, and so on throughout the color spectrum. Now, since Amy is "normal," we say (having discovered the switch) red objects look green to Sid—in the presence of red objects (in otherwise normal circumstances) Sid is presented with +green+ (the phenomenal aspect of his color experience). Moreover, we are to imagine that Sid and Amy adapt to the world in roughly the same way:
Sid stops at red (his +green+) lights, wears green (his +red+) on 17 March, uses his color words as do others, etc. In fact, properly circumscribed, Sid and Army are behaviorally indistinguishable.

Here behaviorism wanes, for on that view inverted spectrum must be metaphysically impossible—by its lights Sid and Amy would, contrary to hypothesis, have to be mentally indistinguishable. This is implausible. We are further to imagine, even given future neurological research, that Sid and Amy are functionally indistinguishable despite the inverting lens—functionalism wanes. For Sid's cornea with its inverted lens plays the same causal role for him that Amy's plays for her—his +green+ is caused in just the way (functionally speaking) her +red+ is caused; the only difference is that Sid's functional characterization has a different physical occupant (the inverted lens) playing the same causal role (or, so goes the objection).

The functionalist's claim that difference in physical realization makes no psychological difference unless there is a functional difference, seems to have been fostered via an analogy with computing machines. Just as machines of different physical structure can execute the same program, brains (as computing machines) can realize the same functional syndrome despite physical (neurological) differences. Where mental talk is talk of functional states, Martians may
realize the same functional states and thus share our mental states (have pains, for instance, correctly attributed of them). But the Martians' physical realizations may differ wildly from ours. To deny the Martian mental states on the basis of a (mere) physical difference would be chauvinistic. Thus a physical difference in the occupant of a causal role would not, it is claimed, amount to a psychological difference (+red+ as opposed to +green+). Sid and Amy, then, are functionally the same though their qualia are inverted. (Notice, however, the questionable assumption that the inverted lens makes no functional difference. This will be addressed below.)

Troubles are alleged to arise for functionalism, then, due to a psychological difference not being matched by a functional difference. So this qualitative difference, a difference that marks a psychological distinction in the way the world is, shows, as Block argues, that the functionalist's proposal for type identification leaves an unexplained residue—the very difficulty Descartes sought to avoid. Sameness of functional syndrome, if the present objection holds, does not serve to identify (all) psychological states, for just as a pain state is distinct from a pleasure state in virtue of the feel, the state of being appeared to +redly+ is distinct from that of being appeared to +greenly+ in virtue of the difference between what it is like
to be in one state rather than another. By now, no doubt, the reader enamored with functionalism is gnawing at the bit. To further examine the functionalist gambit, let us next turn to some functionalist responses to this qualia objection.

**Onward Churchland Science**

Paul and Patricia Churchland, though not themselves functionalists, attempt to meet the above qualia objection by ramming head on the intuition that bolsters it. That intuition—the pro-qualia intuition—can be jostled into service here by considering the question: Can a person be in pain when she does not experience +pain+? Another way of understanding this is by asking if the following is true:

\[ \Box (x \text{ is in pain}) \rightarrow x \text{ is aware of } +\text{pain}+ \]  

Though many prefer this over Kripke's version,

\[ \Box (x \text{ is in pain } \leftrightarrow x \text{ is aware of } +\text{pain}+) \]

I am sympathetic with Kripke's claim for the following reasons. One might be tempted to suppose that the sensation of pain (+pain+) is to be identified with an agent's awareness of +pain+ where that awareness is itself a cognitive act. Since this would lead us into the den of the strong incorrigibility thesis, the claim that pain just is
(essentially) the sensation of pain (+pain+) would falter by no longer having the intuitive plausibility (irresistibility) Kripke claims it has. Yet, recalling Descartes' distinction among three grades of sensation, we might well argue that there is a de re awareness of +pain+ that is not a cognitive act requiring judgment. Rather, such awareness is more like perception, namely, the perception of the "what it is like" feature which is ostensively identified.

In order to (begin to) get a handle on this vexing notion, compare de dicto and de re beliefs. Ernest Sosa provides an apt account that captures the distinction:

Belief de dicto is belief that a certain dictum (or proposition) is true, whereas belief de re is belief about a particular res (or thing) that it has a certain property. (And similarly for knowledge, desire, etc., and far more complicated cases.)

... Suppose that there is some spy Holmes suspects of being a murderer, whereas Watson suspects only that there is some spy who is a murderer... Watson is inclined to believe the proposition that at least one spy is a murderer, but Holmes does more: he suspects about a particular spy that he is a murderer.

Watson's belief (de dicto) can be given the following formal characterization:

(1) $B_w (\exists x) (x$ is a spy and $x$ is a murderer).

And Holmes' belief (de re) the following:
Now I'm not particularly interested in the formal characterizations of these beliefs nor in the formal manner of the characterization of the underlying problems such beliefs generate. Indeed, I am not interested in my (your) belief (de re or de dicto) in any way that can be divorced from my (your) awareness of felt pain. That is, as I shall attempt to make clear, the distinction between awareness de dicto and awareness de re seems more radical (and "far more complicated"). Compare the situation in which Hubert = the chairperson and

(1) Winston believes Hubert is in the office.

Certainly, (1) does not entail

(2) Winston believes the chairperson is in the office,

(nor does (2) entail (1)) for Winston may not know Hubert is the chairperson (may not "know him under that description"). What is even more important, for my purposes, is that (2) can be true even if there is no chairperson. Now, compare the following (where pain = \( N_2 \) realizing \( S_p \)):

(3) Winston is aware of a searing pain,

(4) Winston is aware of his inner physical state, \( N_2 \), that is realizing the appropriate functional state, \( S_p \).
Again, if awareness is de dicto, (3) does not entail (4), nor (4) entail (3). Here it appears there is a defining (identifying) characteristic of pain (of the pain state) of which one could be aware, but of which Winston is not aware, namely, pain is (identified as) physical state \( N_2 \) realizing functional state \( S_p \). This, I maintain, is not plausible vis à vis awareness. Let me explain.

Capturing this de dicto sense of awareness, we would formally render (3) as

(iii) \( A_w (\exists x) (x \text{ is a pain and } x \text{ is searing}) \)

and (4) as

(iv) \( A_w (\exists x) (x \text{ is } N_2 \text{ and } x \text{ realizes } S_p) \).

Thus, it would seem that Winston could be aware of his pain without being aware of his inner state realizing a functional type. It should also follow that Winston can be aware of the latter without being aware of the former even though, given the identity, the former is the latter--where we understand "aware of" as "aware of de dicto." This strains the imagination unless one keeps reminding oneself that Winston is ignorant of the (alleged) identity and we are speaking of awareness de dicto. Even when "believes that" is substituted for "is aware of" the strain to the imagination does not completely vanish. Consider
Winston believes that he is in searing pain,

Winston believes that physical state \( N_2 \) is realizing functional state \( S_p \).

Obviously (5) does not entail (6), nor does (6) entail (5), though this latter is not "clean." That is, if the identity claim is true and if Winston's belief that he is in searing pain is (a) true (belief), then it is true that \( N_2 \) is realizing \( S_p \) even if Winston does not believe the latter. Moreover, if the identity claim is true and if Winston's belief that \( N_2 \) is realizing \( S_p \) is (a) true (belief), then it is true that Winston is in searing pain even if Winston does not believe that he is. Here is where we run into the pro-qualia intuition.

I maintain that the feature by which Winston knows he is in pain, the feature he is aware of and by which he discriminates his being in pain from his not being in pain does not possess (that feature does not possess) any aspects of which someone else could be aware but of which Winston is not aware. To put off the immediate charge of question begging, consider the following:

(7) Lois Lane adores Superman,

(8) Superman exists.

Certainly (7) does not entail (8) (Lois Lane ≠ Margot Kidder and Superman ≠ Christopher Reeves), since the
conditions needed for an entity to be Superman include conditions over and above those it need possess in order to be an object of admiration. That is, for x to be Superman x must have special powers, F, G, H; for Lois Lane to adore x does not entail that anything exemplify these powers. But pains, I maintain, are the sorts of things whose existence is guaranteed by their being objects of awareness. After all, what features would something have to possess to be an object of awareness, a felt pain (+pain+) in addition to those possessed by the object of Winston's awareness?

It will not do to claim that such awareness is always of something, some apparent physical property instance. For to claim, as would Smart, that Winston is aware of something physical, is to claim that Winston takes the feature presented in experience—the "appearance"—to be something physical and external (or he would if he but knew more of his inner workings). Yet, even if Winston were a reductive materialist, he would still be aware of the physical item N₂ realizing Sₚ via its appearance and the appearance need not be physical even if Winston (Smart) takes it to be. Moreover, commonsense and common experience go against the view that such features are physical, for we (each in our own case) are able (indeed, we do it!) to make identifying reference to, to discriminate, again and again, qualitatively
similar objects of awareness whether such objects be construed as intentionally inexistdent or as existing.15

As Robinson remarks with regard to the claim that what experience acquaints one with is what objects that possess that quality are like and not with the quality possessed: "One recognizes an object by recognizing its features and, therefore, knowing what an object with a certain quality is like involves knowing what the quality is. Thus it would be a mistake to say that one is aware of something possessing the quality, but not of the quality itself."16 It is not at all easy to understand what is meant when it is claimed that such features (qualities) are intentionally inexistdent. For one thing, my being aware of, say, an after-image is not happily likened to my being aware of an abstract item existing only in the way in which, e.g., numbers exist.

Again Robinson makes the point:

If S is aware of quality-object F in some area of his visual field, then he cannot also be aware (in that sense of 'aware' in which it makes possible ostensive teaching) of any other comparable quality on the same line of vision. . . . This, I think, gives a clear and relevant sense in which the quality-object of awareness is an empirical, rather than an abstract, entity. It shows that such objects possess what might be called a blocking function. . . .17

(Note, as I will show below, that the sense of "aware" that makes possible ostensive teaching is de re.)
Furthermore, as Robinson goes on to argue, the claim that such objects are non-existent depends on their being intentional and they are "suspiciously dissimilar to other objects." Dissimilar in that, for one thing, the salient fact about intentional objects is that there is no limit to the indeterminateness of the aspect under which an entity might become an object of the relevant activity. Whereas with regard to the objects of awareness, the empirical nature of the phenomenon requires a high degree of determinateness: "I can, of course, say that I am aware of a colour without saying which, but I cannot just be aware of colour simpliciter."18

As we have already seen, if Winston believes Hubert is in the office he does so under some description of Hubert (e.g., "My friend with whom I drank the '47 Petrus"), and that he believes Hubert, under that description, is in the office does not entail that he believe Hubert, under another description (e.g., "The chairperson"), is in the office. Hence, the familiar non-substitutability of intentional contexts; here, then, is another dissimilarity. To be aware of my pain state I must be aware of it by way of some qualitative aspect (+pain+, e.g., searing, throbbing); I cannot be aware of it simply under the description of it as being a physical state realizing a functional syndrome: I cannot be aware of my pain state under that aspect
(description) just by experiencing it. (Under what aspect could someone else be aware of my pain, of which I am not aware?) Thus, since I am aware of my pain state in virtue of its qualitative feature (+pain+) (and under any description I, or any one else, wishes to use), then I am aware of my pain state (under any other description), regardless of description, in only one way—in virtue of its feel or qualitative feature. That is, if it is true that I am aware of my pain state at \( t \) under the description, "whatever it is that happens when [public event]," then I am also aware of my pain state at \( t \) under any other description, e.g., "physical state \( N_2 \) realizing functional state \( S_p \)," because the object of my awareness (+pain+), unlike the object of Winston's belief (or Lois Lane's admiration) is not in any way dependent upon, or tied to descriptions. But if so, then this points out a further dissimilarity between property instances ("quality-objects") one is aware of and intentional objects strictly so-called: non-substitutability fails with regard to the former verb ("is aware of"). Hence, the grammatical features of so-called intentional verbs must, in the case under consideration, give way to extensional idioms, and a deeper look at the phenomenon which is intrinsic to human consciousness is called for.\(^\text{19}\)

I do not (you do not) discriminate or re-identify my (your) pain by noting behavior (overt or covert) or
behavior-cum-public-circumstances. Thus, we should render

(3) Winston is aware of a searing pain,

not as (iii) (de dicto), but as the de re

(v) (\(\exists x\)) (x is a pain and \(A_w\) (x is searing)).

(Note, then, that Winston need not be aware that that which is searing is [what is called] a pain.) Should the materialist wish to identify pain with a physical state, he must also make plausible the further step of claiming that how that pain appears (that is, that that of which Winston is aware, the searing feature, the hurt—that property instance) is nothing but a physical property instance. To do this, as has been alluded to, it must also be claimed that the ostensive learning (teaching) of "pain" must not necessarily involve awareness de re of a non-physical property instance. But in teaching a student, S, the meaning of (the reference of) "pain" so that S will be able to discriminate and re-identify his state as a pain state, the teacher, T, must know that to which "pain" refers, know when that typically occurs and assume (leaning on evolutionary theory perhaps) that S, when in appropriate public circumstances, experiences (is aware of) what felt pain is like. Again, "what felt pain is like" refers to a non-physical property instance. Both T and S, since ostensive teaching is successful, are aware of this what it is like feature ("quality-object").
As Robinson points out, ostensive teaching is even possible when the experience is not veridical; I may teach someone the color pink by knowing that a certain drug they have ingested causes them to be appeared to pinkly. It is even possible for one to non-veridically "see" a quality for which one knows no name and later, upon looking at a color chart, exclaim, "Oh, so that was an experience (an awareness) of what is labeled 'taupe'; I was aware of what the color is like, but not what it is called." A more homely example (of which I am not all that sure works) is to be found in the depth perception tests administered to flyers. It is assumed that a "normal" percipient will see things as they are, viz., A in front of B. But a percipient with a physiological abnormality will not see things this way (as they are). Hence, the degree of precision in what is seen will vary (as the Air Force would say) with fitness. An enlightened tester who knows of the abnormal physiological condition of \( S_A \) will assume that \( S_A \) is appeared to differently than \( S_N \). But now, if \( S_N \) is aware of physical entities (as they really are), certainly what \( S_A \) is aware of is not identical with such physical entities, but how they appear to him. Now this may be "old hat" and this conclusion denied; but though this may be denied, I find little plausibility in the denials as they pertain to awareness and take it that any (plausible) theory must be built around this datum.
At any rate, that awareness is *de dicto*, a cognitive state (requiring some linguistic competence), is far from plausible: awareness of *pain* appears to be *de re*. In this latter sense of awareness one need not know that what one is aware of is a pain (is labelled "pain") in order to be aware of a pain—indeed one need not even entertain the thought that one is in pain. So one could have pains of which one is unaware (*de dicto*) as well as have the mistaken belief that one is in pain—perhaps, it is conceivable, a child ostensively identifies what the word "pain" refers to with the fear quale (+fear*+). Still, however, it can plausibly (more plausibly) be maintained that one is in the pain state if, and only if one has *de re* awareness of *pain*. As is obvious this amounts to the claim that *pain* is essential to being in pain.

Let us turn to Churchland and Churchland. On their view the functionalist need not and should not deny that our sensations have intrinsic properties, and he should agree as well that those properties are the principal means of our introspective discrimination of one kind of sensation from another. What he is committed to denying is that any particular quale is essential to the identification of any particular type of mental state. Initially they may seem to be essential, but reflection will reveal that they do not have and should not be conceded that status.
So Churchland and Churchland see the task as one of determining whether it is the quale or the functional syndrome that enables us to type identify particular psychological states, such as the pain state. In their terminology, which of these alternatives dominates for the purpose of the type identification of such states? Still another way of seeing what is at issue is to recall the (Type-Type) Identity Theorist's claim that we differentiate mental states from among the psychological states according to their physiological kinds. Hence, the latter view, like the functionalist's, must deny that qualia, construed as mental items (non-physical property instances), are essential to this task (with regard to both first- and third-person discrimination). For the functionalist, which among an organism's (system's) physiological (physical) state-tokens are the mental ones of the various mental types, depends on what role those tokens play in the organism's (system's) functional profile. Thus, should Churchland and Churchland develop a strong case for the claim that a particular quale (e.g., +pain+) is not essential for being in pain (thereby overturning what has been argued above), then the fact that inverted qualia are possible does not sully the functionalist picture, for it can then be maintained that two individuals with inverted qualia are nonetheless in the same psychological state. Notice that the possibility of inverted qualia poses
no special difficulty for the (Type) Identity Theorist, for cases of inverted qualia, should they occur, would be accounted for by noting that there must be a difference in relevant physiological state-tokens if there is a difference in mental state (and vice versa). (This strong empirical constraint is what troubles the Type-Identity claim, not the possibility of inverted qualia.) Of course, as we shall see, to claim that inverted qualia present no special problem for the Identity Theorist is not to say that qualia present no difficulty.

With regard to Churchland and Churchland, we shall discover that their notion of domination leaves it open that, although the state of being in pain may not always be accompanied by +pain+, it may always be accompanied by some qualitative feature. It may also, of course, sometimes be accompanied by +pain+. For them, as well as the Identity Theorist, we must note that if this accompanying qualitative feature is not a causally operative factor in the ensuing behavior, we appear stuck with an ontological mule. Thus, if a particular quale is not essential for my identifying my psychological state as a pain state, thus undercutting (at least part of) the intuition supporting the qualia objection, it would seem to follow that I can become aware that I am in pain and engage in avoidance behavior though there is no +pain+ (unless [de_re] awareness if, and only if +pain+.)
But this, it has been claimed,\textsuperscript{22} is problematic for the evolutionary account of the emergence of +pain+. That is, if we think of +pain+ as causally efficacious in generating avoidance behavior, then propositional awareness of pain without +pain+ may be thought unacceptable, from an evolutionary perspective, because it would not be clear how such beliefs could be generated (caused) without +pains+. This may be thought especially so since there is good reason to believe that the person who, as a matter of fact, is congenitally unable to feel +pain+ is at a distinct biological disadvantage. Consider the case of Miss C., whose

\ldots medical history revealed a long record of hospitalization for injuries and illnesses connected with inability to feel pain. At the age of 3, for example, she suffered third-degree burns from kneeling on a hot radiator while looking through a window at children playing in the street. Once she incurred severe sunburn on the backs of her legs which went unnoticed until the area became infected. The tip of her tongue was deformed from biting it in temper tantrums as a child, and she had multiple scars on her hands, legs and feet. She stated that she had never experienced headache, earache, toothache, stomach ache or menstrual pain.

Neurological examination confirmed her insensitivity to pain. She had no corneal reflex, though she could feel a very light touch on the conjunctiva of her eye. Suborbital pressure did not hurt her. Insertion of a stick on the nostrils produced no withdrawal reaction, nor did severe pinching of the Achilles tendon. Histamine injection failed to give her a headache. Pinprick was not more painful to her than finger touch.
A hypodermic needle was felt penetrating the layers of skin, but without discomfort.

We do not know what neural defect accounts for Miss C.'s remarkable analgesia. Histological examination of peripheral cutaneous innervation revealed no abnormalities. Nor did neuropathological study of her central nervous system at autopsy some years later. It is interesting that while Miss C. was curious about the nature of the feeling accompanying the pronounced pain reactions she observed in others, she apparently had no concept of this herself. For example she once expressed to the examiner her belief that as she got older she was becoming more sensitive to pain. But upon further questioning it turned out she meant by this only that with experience she was becoming more and more adept at avoiding injury in the absence of pain warnings.

Although it might be claimed that Miss C. (who died at or about 26) was at a biological disadvantage not because she was unable to feel pain, but because she had inappropriate neural mechanisms (she was "wired" wrongly), this claim appears to falter (so Puccetti might argue) over the import of the "because" involved. If what was causally responsible for her inabilities was merely a defect in neural structure, then it would seem that it is neural items alone that make up (realize) the causal chain from stimulus to avoidance behavior. (It it is argued that Mackie's nod in the direction of epiphenomenalism can account for our beliefs that we are in pain without maintaining that +pains+ are causally efficacious, then we are still left with
ontological worries [see ch. 1].) However, since most of us do feel +pain+, this implies that such feelings are (if anything) purely gratuitous occurrences we are better off without, for having the appropriate neural hookups should be enough for avoidance, etc., behavior—there would be no biological disadvantage to not feeling +pain+. So if +pains+ are not needed in addition to neural mechanisms, why are there +pains+? Since most of us do (luckily!) feel +pain+, Puccetti, as we have seen, finds it difficult to grasp how an evolutionary account of their occurrence (emergence, supervenience) can be given unless they are construed as items playing an operative role in the causal path. Again, should we follow Mackie's suggestion and maintain that felt pains are epiphenomenal, we remain consistent with evolutionary considerations (as well as the data of introspection), but at the price of endorsing a non-materialist view. As I have said, I find the latter tack more acceptable than Puccetti's interactionism.

Saying that one feels pain if, and only if a certain functional syndrome is realized, adds little to aid the understanding, for then we are back to the question of the nature of the realizers and the ontological question of what +pains+ are. Thus the claim that there is +pain+ but that it is not essential to natural selection, in the sense that it (construed as a quality-object) does not serve as
causal factor, appears to leave an unexplained qualitative residue—epiphenomenalism threatens. On the other hand, if +pains+ do play a causal role, interactionism threatens.

My present target, however, is Churchland and Churchland's attempt to squelch our (at least my) intuitions concerning the importance of qualia for type identification of (at least some) psychological states. As I shall argue, a major difficulty with their claim is their tendency to view type identification too narrowly by focusing only on the asymmetrical ascription of pain to another (for emphasis, an other), leaving aside the more interesting problems of what is involved in recognizing our own psychological state as a pain state. I turn now to their argument.

Opposed to our homespun intuition, they argue, is another story with "intuitive appeal" of its own. Their suggestion is that my pain/sensation-of-green may have a qualitative character "rather different" from yours, and these internal differences among the "same" psychological states are not only possible but perhaps not very unusual. Thus, "the functionalist should concede the juggled qualia, while continuing to reckon type identity in accordance with functional syndrome."27 (Again, note that these juggled qualia remain as unexplained features of psychological states, unless elimination or reduction is offered.) Although their ploy is to bolster this "intuitive appeal" by
scotching its competitor, I prefer to first query as to what intuitions they believe answer to their claim; in so doing we shall obtain a bit deeper insight into the strength of the intuition they are out to overthrow.

Certainly it does appear that my +pain+ may differ from yours even though we are in roughly the same situation vis-à-vis the causal story. After all, my own pain states do not all have the same (identical) qualitative feel; there is a quite noticeable variation in feel between a searing burn and a splitting headache—I can distinguish the two feels though they are both instances of pain (+pain+). Moreover, a bit of reflection lends considerable support to the (at least) prima facie conviction that we mark this distinction by noting the difference between what it is like to be in one state rather than the other. It hardly follows, as Churchland and Churchland claim, that the wide variety of qualia lumped together under the heading of pain makes it "evident that what unites sensations of such diverse characters is the similarity in their functional roles." Though my headache and yours no doubt have a similar feel, I would not insist the feel must be the same (identical?) for us both to be in pain—there may only be an identifiable likeness just as when there is an identifiable likeness which I note in my own case. But I would deny that my feel and yours could be radically different (pain/pleasure quale
inversion) and still both be instances of pain. If what it felt like for you to be in pain were what it felt like for me to experience pleasure, then despite the fact that our functional syndromes were alike we would not be experiencing sensations of the same kind and would not be in the same (type of) psychological state. Surely pain/pleasure quale inversion (yet same functional profile) is anomalous, a mere logical possibility. It is better to imagine only that the qualitative character of pain in X is different from its qualitative character in Y. Nonetheless, were the admitted difference great enough (admittedly vague), then despite similar functional profiles, X (say) would and Y would not be in pain. It may even be that a difference in qualitative character must (nomologically) be accompanied by a functional difference—still this is not to say that qualitative features are not and functional features are essential. Rather than similarity in functional roles, it is similarity in qualitative character that enables me to identify and distinguish a searing from a throbbing pain sensation (of mine) and either from a tickle. When I touch a red-hot iron, functional profile plays no role in my determining my psychological state; the urge I feel to withdraw my hand is always accompanied by (or due to) +pain+ (albeit not identical with it). That is, I need know nothing of functional profile to identify my psychological state as a pain state, although
another must take the (sincere) report of a pain feeler before he can label a functional profile as a pain-functional profile and thus claim, on that basis, that I am in pain. As argued in Chapter 1, knowing all functional/physical facts would not amount to knowing all there is to know about an organism in pain. This latter denial lends strong support to the pro-qualia intuition, and the above reasoning is sufficient for concluding that there is no stronger intuition, of a functional nature, which accounts for the variation among the qualitative characters of our pains.

One might be inclined toward this alternative, however, when one considers type identifying an other's psychological states. How is it that I can identify Morgan's psychological state as a pain state? To slip the "other minds" conundrum, it may be wise to simply opt for the functional syndrome alternative as the best that can be accomplished while avoiding the naive peripheralism of the behaviorist. Nevertheless, the asymmetry here is striking, for if I were to discover that Morgan, despite his pain-functional profile, did not feel pain (+pain+) like that which I feel, I would deny that he was in pain (despite whatever puzzlement might befall me given his behavior). So even though it may be plausible to argue that identifying an instance of pain in an other requires use of the functional syndrome tack, that hardly presents sufficient reason for thinking that a
particular functional characterization is essential to pain or, alternatively, that it not be essential that a particular quale be present. To further capture this distinction and see just what use the functional syndrome line may have with regard to language learning (concept formation), consider a remark of Wittgenstein's:

> Or is it like this: the word 'red' means something known to everyone; and in addition, for each person, it means something known only to him? (Or perhaps rather: it refers to something known only to him.)

The distinction Wittgenstein points to in this rhetorical question marks the asymmetry mentioned above as well as indicating the nature of the connection between felt inner sensation—"private" in the sense of being a psychological entity picked out via "private" ostension—and functional syndrome—a completely public account/description not requiring ostension—with regard to language learning. As Cornman points out, this also enables one to avoid (at least one version of) the Private Language Argument against the notion that psychological sentences (first-person reports) can be used to make identifying reference to private psychological entities. To see this let us consider the above quote replacing "red" with "pain".

What the word "pain" means, in the sense of being known by everyone (accomplished in sensation-language use), does
depend on a public occasion (or description of such an occasion) which cues us as to when "pain" is to be used to attribute a mental state to another. We also learn, initially at least, when to attribute pain to ourselves, that is, when to correctly apply the term "pain" to our own state, in the same manner. Where we let "know₁" refer to knowing how to correctly to use the language, then we come to know₁ when it is correct to attribute pain to others—and initially to ourselves—when we learn the public criteria/description backing the attribution. Yet an asymmetry quickly develops, for when we learn what "pain" refers to, we learn this ostensively by noting the "what it is like" feature accompanying the publicly observable cues used to correctly attribute pains to others. Indeed, it is this latter act of ostension which enables us to distinguish the pain state (i.e., +pain+) from the circumstances which accompany it—and we do do that. Thus, when we report that we are in pain, we come to know₂ something "known only to us". For when we know₂ what "pain" refers to, we are able to identify our own psychological state as a pain state without dependence upon public criteria (or public description)—even though knowing₁ that we are in pain does require an initial public occasion/description. Hence, in the case of others it is public occasions (wincing, crying out, etc.) or functional descriptions that back our attribution of
mental states. In our own case, however, it is the "what it is like" feature which backs our reports (self attributions). Since knowing how to use the term "pain" does not require identifying reference, it does not follow that only one individual can know the rules for the use of "pain" (although the referent of "pain" is picked out via ostension) and hence, the Private language Argument is thwarted.33

Intuitively, then, the way we identify (in our own case) a particular sensation as a pain is by likeness of qualitative feel—no such feel, no pain despite the fact that others may be warranted in using "pain" to identify our state. Contra Churchland and Churchland, then, the functional syndrome line does not have a "certain intuitive appeal of its own"—or at least not one that outweighs the pro-qualia intuition. By focusing on another (or pretending to be anesthetized), they simply ignore this intuition. Still, however, it may be theoretically preferable34 should the pro-qualia gambit be shown to be grievously in error. Let us return, then, to their tactic.

One major source of support for our pro-qualia intuitions is due, claim Churchland and Churchland, to our plebeian folk psychologizing where we have considerable success explaining and (especially) predicting human behavior. Since talk of psychological states figures prominently in such endeavors, we are tempted to view our
successes as due to, e.g., pains all having quite similar intrinsic properties. As Putman\textsuperscript{36} has forcefully argued, science has shown us there is an intrinsic nature essential to every instance of, say, water. Similarly we conclude, claim Churchland and Churchland, that there is an intrinsic nature essential to every instance of pain. Now, although this conclusion sounds right, I'm not at all sure I find it appealing because pain is like water in that we attribute essential properties to it on the basis of explanatory/predictive success.\textsuperscript{37} Indeed, much more needs to be said if my reasons for believing that a particular type of quale is the intrinsic property essential to being in pain are to be captured. We move, then, in that direction.

Our pro-qualia intuitions are, it seems, further encouraged when we attempt to complete the analogy: just as the intrinsic nature of water is $\text{H}_2\text{O}$, the intrinsic nature of pain is ____. It is by noting the deliverances of introspection that we fill in the blank. As we have already seen, type identifying our own pain states, as contrasted with those of an other, feeds off our introspective abilities and gives rise to an asymmetry in our epistemic position vis-à-vis that other's psychological states--I cannot know Morgan is in pain as I can know I am in pain. Therefore we conclude, they argue, that the feature essential to being in pain is the qualitative feature--if one does not feel
pain, one is not experiencing pain, is not in the pain state—
not really.

As intuitive as the above reckoning is, Churchland and
Churchland view it as resting on a dual mistake: psychol-
ological states do not amount to natural kinds and introspec-
tive "facts" cannot serve to ground pro-qualia intuitions.
Reliance on an unexplicated notion of physical science bears
the burden of their argument here. They begin by noting
that:

Given the physiological and chemical variety
we find in the nervous systems of the many
animals that feel pain, it appears very
unlikely that their pain states have a common
physical nature underlying their common
functional nature. . . . It remains possible
that they all have some intrinsic non-
physical nature in common, but dualism is
profoundly implausible on sheer evolutionary
grounds. 

As they see it, if pains are natural kinds, then either
there is an objective intrinsic physical nature all pains
have in common or a non-objective intrinsic non-physical
nature. But it is the apparent falsity of the first disjunct
that helped motivate the move to functionalism in the first
place, while the second is to be avoided because "[t]he
evolutionary process just is the diachronic articulation of
matter and energy." And with this heavy handed move, the
juggernaut of science thuds on to the conclusion that "the
empirical presumption against natural kind status for
psychological states is substantial."\(^{40}\) So, given that "pain" is not a natural kind term, pain has no essential intrinsic nature and hence, no quale is essential contra the pro-qualia intuition.

As an argument against natural kind status for psychological states, the above is a gaffe convincing only to those already enamored with a totally unexplicated picture of future science. For one thing, it may be that there is a third alternative, namely, that there are objective functional (non-physical) essences which all pains "share."\(^{41}\) But more of this possibility in Chapter 4; here I shall concentrate on the sweeping claim that "dualism is profoundly implausible on sheer evolutionary grounds." As Wilfrid Sellars would complain of this glib treatment of serious puzzles concerning the nature and status of sensory consciousness:

\[\ldots\text{what I find objectionable in the views of many of my tough-minded colleagues is not their "extreme scientific realism," but rather their failure to pay serious attention to the problem of specifying the conditions which an adequate scientific account of human behavior must meet.}\ldots\]

This failure leads to a reliance on overly simply and inadequate paradigms of what will count as a "scientific object" or "bodily state" or "neurophysiological process" in this anticipated scientific account.\(^{42}\)

Per my earlier discussion in Chapter 1, evolutionarily considerations may fail to show epiphenomenalism implausible
and fail to supply us with an adequate materialistic account of how +pain+, as an emergent/supervenient property instance, is to be accorded nonmysterious ontological footing (while still accounting for the phenomenological data). Indeed, such considerations seem only to raise the ontological worries traditionally associated with dualism. So it is not at all clear how evolutionary theory renders dualism (of one sort or another) implausible, for if the causal schema leading to avoidance behavior has only neurological (better, physical) items occupying the relevant causal roles, epiphenomenalism dangles an unexplained group of (non-physical) properties before us. Talk about causal roles *simpliciter*, with no mention of the items filling those causal roles (pure abstract talk of functional properties), likewise fails to indicate how +pain+ is to be made consistent with materialism. Moreover, if Churchland and Churchland are maintaining that the evolutionary process is a part of a systematic materialism in virtue of being explained via conservation laws, thermodynamics, etc., then they are merely issuing a check to be drawn against the aforementioned anticipated (and unexplicated) scientific account. After all, there may be lawlike regularities (+pain+ whenever a particular functional syndrome is realized) without there being an ontological resolution. Thus the emergence of +pain+, as an operative causal factor—a mover of things—
seems to require some sort of interaction between nonphysical (mental) properties and neurological (physical) properties. Moreover, if the causal challenge can be met, we still have (unless reduction or elimination) epiphenomena to worry about. Dualism, then, is not shown "profoundly implausible" on evolutionary grounds alone unless the very qualitative feels we use to recognize some of our own psychological states can be accounted for within the theoretical structure of contemporary science: vis-à-vis the mind-body problem, qualitative features remain an enigma. (For an unconvincing argument that we will learn to perceive the world directly in terms of the conceptual framework of (future) physical theory (a "conceptual revolution"), see Paul Churchland's Scientific Realism and the Plasticity of Mind. Still, it remains that a +pain+ by any other name is a +pain+.)

So far, then, the above dilemma appears to crumble because (i) a possible third fork is not considered, and (ii) given that there is no argument against (nor explanation of) the existence of qualia, the real possibility of the latter having the status of epiphenomena or involving us in interactionist difficulties remains, thereby undercutting the claim that dualism is shown profoundly implausible (it is, of course, profoundly mysterious). Failing to show that qualia do not exist (or otherwise accounting for
then), leaves it wide open, as far as the above argument goes, that the pro-qualia intuition is sound—the qualia are "there" and if they weren't a particular psychological state would not be, e.g., a pain state. Perhaps, then, the real argument is against the deliverances of introspection.

Science and Introspection

Consider next an analogy appealed to by Churchland and Churchland. Our usual way of visually recognizing tigers is by noticing their black and yellow stripes; nonetheless such stripes are not an essential feature of tigerhood. As I argue, such analogies—dealing with non-introspective discrimination—play a prominent role in motivating the "intuition" lying behind Churchland and Churchland's claim that scientific progress will lead us to abandon phenomenal properties as the most reliable (indeed, essential) indicators that our psychological state is a, e.g., pain state. Thus they conclude: "why should the qualia of our familiar psychological states be thought any different [than the tiger's stripes]?" It is their proposed answer to this key question with which I am unhappy. We must be careful here of slipping too easily toward this beguiling claim. I prefer to limit discussion to pain, for some psychological states, for instance, jealousy, envy, anger and love, while having qualitative aspects, do not depend on our
introspective abilities for their accurate recognition (one need not be a Freud scholar to realize that we are not only often mistaken in attributing such states to ourselves, but often manage to deceive ourselves). What this points out is that the "teeming chaos of our inner lives" consists of qualia we learn to recognize, to name, because some of that chaos more readily stands out, for example, +pain+. Thus, although it is true that I must learn to use the word "pain", it is a sizable leap to the conclusion that:

... since the taxonomy of observational qualia constructed by the questing child follows the discovered taxonomy of states as determined by interesting causal roles, it is evident that sameness of functional role dominates over the difference in qualitative character, so far as the type identity of psychological states is concerned.46

Again care is needed; in one sense (at one time) functional role does dominate, for certain purposes, by providing clues as to when to apply words (ascribe states) to ourselves and others. Though type identity is tied to language learning at one time, a look beyond the original language learning of the questing child indicates that an asymmetry quickly develops. Consider the sentences:

(1) Morgan is in pain, poor fellow.
(2) I am in pain, damn it!
With regard to (1), considered in the appropriate context, my ascription of pain to Morgan will typically be the result of my witnessing a public stimulus (e.g., a brick lands on his toe) followed by Morgan's pain-behavior. Concerning (2), construed as a report, I (it would seem) first learned to use "pain" by noticing happenings involving bodily injury and hearing others ascribe pain, hurt, etc.; thus the referential force of "pain" is closely tied initially to public criteria ("stage setting"). Eventually, however, the nonpublic stimulus gains control and I attribute pain (answer "yes" to "pain?") to myself only if +pain+. (Moreover, I am sympathetic (say) toward another exhibiting pain-behavior only because I can imagine myself feeling pain.) Hence, while functional role may have dominated in my original language learning and still dominate qua sentences like (1) (ascriptions to others), certainly the law-like occurrence of the nonpublic stimulus, +pain+, whenever "pain" was accurately applied to me, gave rise, qua sentences such as (2), to a quale dominated recognition of the accuracy of the ascription. Failure to make this transition, to distinguish between +pain+ and its accompanying (public) circumstances, would leave one unable to understand the difference between real pain and pain in a performance, a plight analogous to the congenital non-pain feeler. The following claim, then, brings to a head the conflict of "intuitions" that underlie
Churchland and Churchland's (anti-qualia) position:

AQ That the qualitative character Q of a psychological state S should serve as the standard ground of S's introspective discrimination is entirely consistent with Q's being a non-essential feature of S.48

Before attempting to account for what plausibility AQ has, let us recognize that there are two ways of understanding the current discussion. The weak way is to take the pro-qualia side as maintaining only that the pro-qualia intuition is not refuted by functionalism. That is, it remains as a competitor with initial plausibility. We may then go on to wonder why it should be preferred if AQ also has plausibility. On this reading it strikes no direct blow at the anti-qualia position, but only holds its own. A stronger way of viewing matters, however, is akin to Kripke's view49 that the pain state has as its essential property the way it feels. We can thus render the pro-qualia conclusion thus: necessarily A is in the pain state iff A is aware (de re) of +pain+. This latter view appears incompatible with AQ. Now, despite all the intricacy of Kripke's arguments and their relevance to other matters, in the present context his case rests upon intuition. (Albeit intuition, I maintain, that is backed by arguments alluded to above and expanded upon in chapters 5 and 6.) But this observation does not denigrate his view since, as I see it,
the pro-qualia intuition sets the conditions which any adequate explanatory framework must satisfy if, in Sellars' terminology, we are to successfully make the conceptual journey from the Manifest to the Scientific Image conception of persons. Functionalism (à la AQ) can be seen as vying for the position of intermediate stage of explanation between these two Images. Since functionalism has many enticing features, it deserves to be put through the market place, for (to update Mill) the best theory will be naturally selected. So, after presenting what I take to be the motivation for AQ, I shall argue that in the case of pain (at least), the pro-qualia intuition (Kripke's version) is to be preferred--the "conceptual revolution" is not yet at hand.

Consider, then, the claim: necessarily W is water iff W is H₂O. Without making use of (or worrying about) possible world apparatus, let us make a few observations that tie this claim to the earlier analogy of the tiger and its stripes. At one time, certainly, the standard and only ground for the discrimination of water was thought to be via such macro-properties as are readily observable by the unaided senses: coolness, tastelessness, odorlessness, wetness, thirstquenchingness, etc. Now, however, water is more reliably and scientifically discriminated by micro-properties, namely, H₂O. What were once, we can imagine, thought to be essential properties of water because they
served as the standard ground for our discrimination of it, nonetheless turned out not to be essential features. As far as I can tell, it is by analogy with cases such as this (and the tiger case) that AQ is thought applicable in the case of pain. Hence, AQ is thought plausible because the following is plausible (if not true);

AQ That the observable feature O of a physical state W should serve as the standard ground of W's discrimination is entirely consistent with O's being a non-essential feature of W.

Where "O" stands for any one (or cluster of) the above listed observable properties once serving as the standard ground for the discrimination of water (W), AQ has, given the Putnam-Kripke line on natural kinds, plausibility. We can, after all, plausibly imagine water (= H₂O) appearing differently in, say, a different atmosphere. Despite this (here the analogy begins to weaken) I cannot plausibly imagine a pain feeling radically different (at the extreme, pain/pleasure quale inversion) while remaining a pain. But the above analogical reasoning would have us claim that just as the observable properties of water are not (as science has shown) essential, neither are the phenomenal properties by which we discriminate pains from other psychological states. Strictly we are to infer, I take it, that science will show such phenomenal properties to be non-essential—
for it has not yet shown this to be the case. Since a particular psychological state at \( t \) has yet been shown to be (or to be plausibly construed as) the realization, by a physical property instance, of a functional state, it is difficult to see how AQ inherits plausibility from AO.

The analogy suffers, of course, because of the difference in the manner of discrimination involved: AQ involves introspective discrimination (awareness \textit{de re}), AO publicly observable discrimination. When I discriminate my state as a pain state I do so via an awareness of what pain is like; it is the what it is like feature (+pain+) which enables me to distinguish my pain state from another psychological state. When I claim to discriminate Morgan's psychological state as a pain state (when I predicate "pain" of him) I do so via publicly observable features (correlated with, what I take to be, his felt pain). While AO may be relevant to the observable features which serve as the standard ground for my discrimination of Morgan's psychological state, it is only because the essential feature of Morgan's pain state, what it feels like for him, is plausibly thought to accompany such features. I may be wrong, these publicly observable features may indeed be non-essential features of his state. It is the difference between first- and third-person discrimination that blocks AQ's inheriting plausibility from AO. Compare: I can be aware of a sort of stuff
I take to be water, and yet be wrong. But when I am aware of +pain+, although I may wrongly describe my state (I don't know the language, I made an (uncorrected) error when first learning "pain"), that I am aware of +pain+ is not something I can be wrong about. Compare: I am aware of an after-image; I may wrongly interpret it as an item in (physical) reality. Still, it cannot plausibly be claimed that I am wrong in claiming to be aware of a quality-object, call it what you may. If anyone is guilty of question begging here, I believe it is Churchland and Churchland. After all, one need not claim that it is logically impossible for AQ to be true (although . . .), but only that there is no plausible reason for thinking it is true. What Churchland and Churchland must claim, here, is that the following instantiation of AQ is plausible because AO is plausible:

\[ \text{AQ}^+ \quad \text{That the qualitative character +pain+ of the pain state S should serve as the standard ground of S's introspective discrimination is entirely consistent with +pain+ being a non-essential feature of S.} \]

\[ \text{AQ}^+, \text{however, does not inherit plausibility from AO unless psychological states are assumed to be just like physical states. But this is, of course, up for grabs unless one is to beg the whole question at issue: qualia are a prima facie counter-example to AQ+. Until there is an independent argument to show psychological states are just} \]
like physical states, the difference between AQ+ and an instantiation of AO is much more severe than the difference between apples and oranges.

At any rate, the pro-qualia intuition underpinning the inverted qualia objection, and in which the immediacy of the above visceral appeal to our acquaintance with +pain+ finds a plausible home, does not succumb to Churchland and Churchland's frontal attack. This serves to indicate just how deeply rooted the pro-qualia intuition and how difficult the averruncation facing the functionalist. We turn next to another attempt to meet the inverted qualia objection.
3. ON LEAVING NOTHING OUT—THE FATA MORGANA TACK

I see no other way to force this issue into the open than to defend an admittedly counterintuitive theory and await enlightenment on what, exactly, I have left out. I have, actually, done this (or roughly this) in the second half of Content and Consciousness, but many generous-minded readers of that book have found ways of softening my apparent excesses and saving me from my own drastically drawn view. In other matters I admit to needing all the exegetical charity I can get, but in this instance I must insist on being taken at my word and shown exactly what is wrong with what I say.¹

"Since they are not, there is evidently nothing to worry about, nothing to leave out. No qualia to be inverted." So might one be tempted to summarize D. C. Dennett's "admittedly counterintuitive theory." Finding the genesis of such a view takes us to the nub of the Quinean explanationist gambit (see below) and demonstrates that the charge of reductio ad absurdum, while easy to make, takes work to make stick. Nonetheless, to be told that there are no pains plausibly invites the charge; a charge the tenor of which I shall attempt to make stick—we shall see that Dennett leaves something out.

Playing Our Game

Consider the question raised earlier. How is it that we are able to discriminate the searing burn from the splitting
headache and either from a tickle? We can, I take it, recognize pain sensations—we can say that the sensation we had today was like the one we had yesterday; that is to say, we are aware of something of a more or less determinate and recognizable nature. To respond that there is an identifiable likeness that all pains share is rightly seen by Dennett as amounting to the claim that "the person does not do anything in order to distinguish pains; he just distinguishes them." For Dennet, an explanation in which a person's motives, desires, beliefs, sensations, etc., figure in the explanans which accounts for the person's actions (as opposed to mere bodily movements) is but one level of explanation. Consequently, Dennett distinguishes between the personal level of explanation and the sub-personal level. The former involves taking the intentional stance, where talk is of persons and their (in this instance) sensations, while the latter is the level of realization, where talk is of (for the most part) neurophysiological items or more abstract functional characterizations. Taking the intentional stance, thus characterizing humans (and other physical systems) in intentional terms, is legitimized by its predictive utility. Mature psychological explanation, however, occurs at the sub-personal level. So his point is that, concerning our discriminating ability, there is nothing more to be said from this [personal]
stance, but from another stance an explanation can be given of this primitive ability we have.\(^6\)

At this point we might begin to worry that the minacious possibility of epiphenomenalism/interactionism is already threatening, for when Dennett asks how we so discriminate, he seeks a causal (mechanical) or functional story— one in which introspection has no privileged position. However, if the causal/functional account of the pain network simply tells us how the quale comes about—not what it is and how it serves as causal factor—we are still left with an ontological quandary (unless, I suppose, one is a committed dualist). Unlike Churchland and Churchland, Dennett appears sensitive to this possibility:

Pains are the identity theorists' most plausible candidates for brain processes, but also in other theories the most compelling examples of 'emergent' qualities or 'epiphenomena'. . . . The question before us now is whether pain is something (some thing) in addition to the physical operations of the pain-network.\(^7\)

Again, his recognition of this difficulty surfaces in a later discussion of Sellars' greatest worry, color:

The Lockean distinction between primary and secondary qualities has remained compelling . . . and it tends to lead to the view that the primary qualities . . . somehow work to produce in our consciousness the secondary qualities . . . and that
these qualities are on an ontological footing with the primary qualities but somehow insusceptible to analysis and explanation within the physical sciences. They are real, but 'emergent'... 8

Since these worries ape our earlier setting of the problem, we see that emphasis here is not on the identity conditions for pains, but on the ontological puzzle of what kind of things they could be—Dennett is playing our game. For if they are emergent, in the nasty sense of non-physical qualities so far inexplicable within the scientific framework, then the pro-qualia intuition remains indomitable in the face of the contemporary scientific juggernaut: we just do recognize pains from other sensations, +pains+ are "there" to be so recognized and abstract talk of functional role provides no explanation of what they are. What, then, is Dennett's way with qualia in general and pains in particular.

Arguing for Nothing

Dennett's staddle is his unswerving desire to avoid multiplying entities beyond necessity. 9 Although this slogan is often heard, seldom does it figure in a more interesting argument in support of so staggering a conclusion. Even so, an interesting argument does not, of itself, reduce entities. In a sense, then, our question centers on what entities are necessary for an adequate
explanation of the objects of nature—Descartes' concern remains our concern. As I read Dennett, his argument comes in tandem steps: step one presents a negative argument to the effect that there are no +pains+, while step two attempts to appease our intuitions (and commonsense) by offering a positive account of that to which we do have (some form of) privileged access, thus accounting for the deliverances of introspection. Since step one underpins his argument, I shall concentrate on it, holding criticism until it and a sketch of step two have been elaborated. Consider, then step one:

P₁ If the pain quale is to do any work in a theory, it must be identified.

P₂ To be identified a quale must be described or ostended.

P₃ But description presupposes analysis and, in the case of pain, analysis presupposes personal activity.

P₄ As we have seen, pain discrimination occurs without personal activity.

P₅ Hence, no description of a discriminated quality is possible.

P₆ Ostending the pain quale, however, cannot be separated from ostending the act of discrimination.

P₇ So, our only clue to the localization (in space and time) of the pain quale is the very act of discriminating it.
C Apart from our ability to discriminate the painful sensation, insisting that there is the pain quale "is thus insisting on an unintelligible extra something."12

A bit of elaboration is needed to complete step one, for it is a rather pawky move that takes us to a new level of explanation where it is concluded that there are no pains. Since at the personal level of explanation +pain+ is an unanalysable quality, "[i]f, in our attempt to build an explanatory bridge between sensation and action here, we invoke the appreciation of an unanalysable quality of painfulness, we are forced to choose between two non-explanations."13 Dennett's claim is that either the awfulness of pain (the reason why we avoid it) is a contingent fact admitting of no explanation because +pain+ is unanalysable, or necessarily (by definition) pain is abhorrent and, once again, no explanation is possible since the "because" of causal explanation is not applicable.14 And so, if we are to avoid a premature end to (causal) explanation, we must abandon the personal stance and embrace a sub-personal level account. Notice that Dennett is here squarely faced with the ontological/causal mystery that we have laid out in Chapter 1. When he asks what it is about painfulness that prompts us to avoid it—withdraw our hand—or attempt to eliminate it—apply salve—his answer is that we cannot allow +pain+ to be
any thing which causes (prompts) our bodily movements, because causal efficacy is only intelligible at the "sub-personal level of brains and events in the nervous system." Dennett's view on these matters, then, is in accord with our earlier diagrams of the "causal path"; thus the claim that we will be faced with a non-explanation is the claim that causal explanation would be rendered unintelligible if +pains+ are recognized as causal factors instead of physical afferent-efferent connections in the nervous system. By ruling +pains+ out of court, he also seeks to avoid supervenient/emergent puzzles. To the worry that the sub-personal stance would leave something out and not be an account of real pain-behavior, Dennett's response takes us to the quick of his eliminativist proclivities. Consider his query:

Could any sense be made of the supposition that a person might hit his thumb with a hammer and be suddenly and overwhelmingly compelled to drop the hammer, suck the thumb, dance about, shriek, moan, cry, etc., and yet still not be experiencing pain?!

Even though we are to imagine a non-actor compelled to so behave, the unexplicated modal "could" permits two readings which engender two opposing answers. In one sense I can make "sense" of such a situation by relying on my pro-qualia intuitions while conceiving of a case of pain/pleasure quale inversion. Still, holding this (what would amount to
a possible worlds' parable) aside, answering "no" would seem to call for my admitting that such behavior is (at least) generally accompanied by +pain+, which is the reason for (cause of) the shrieking, moaning, etc. So what is Dennett's point? His Wittgensteinian move indicates, as I read him, that he wishes to identify the behavior and (given his views in *Brainstorms*, its obvious functional causes as all there is (if anything) to the pain state, thus denying any qualitative feature. Consider, as further support for C, his claim with an attached footnoted remark by Wittgenstein:

Positing some horrible . . . quality or phenomena to accompany such a compelled performance is entirely gratuitous.

"And yet you again and again reach the conclusion that the sensation itself is a nothing." Not at all! It is not a something, but not a nothing either. The conclusion was only that a nothing would serve just as well as a something about which nothing could be said.17

Although, as I shall argue, the give and take of this aphorism combined with the positing talk breeds confusion, let us continue with the argument.18 We are, then, to conclude that there is no such thing as +pain+ because, as in C, its being unanalysable makes it unfit for inclusion in an explanatory theory.19 Makes what unfit? The pain quale. What pain quale? The one "we" believe is "there" but really
isn't (fata morgana) since, à la Quine, it is eliminated from "our theory" (hence "pain" is not quantified over) due to its not serving "our" explanatory purposes. While in an instrumental sense our belief, at the personal level, that there are +pains+ is "true" (it figures in successful predictions), when it comes to offering a mature scientific (robust causal) explanation of human experience, we discover that no such things are needed in our account. Although we are almost ready for the beginning of Dennett's positive account of pain, one puzzling move needs inurement.

Since Dennett, like Sellars, seeks to analyze "the relationship between our vision of ourselves as responsible, free, rational agents, and our vision of ourselves as complex parts of the physical world of science," it would seem that the two levels of explanation must be related. And yet, as Dennett remarks, "if there is to be any relation between pains and neural impulses, they will have to be related by either identity or non-identity. . . ." To avoid this fork we must, claims Dennett, recognize that "different categories are no better than different Cartesian substances unless they are construed as different ontological categories and only one category of terms is referential." As the problem has been set, it is causal worries that form the basis of concern here. Thus, in order "to prevent the contamination of the physical story with unanalysable
qualities of 'emergent' phenomena," we are told that "it is in one sense true that there is no relation between pains and neural impulses, because there are no pains; "pain" does not refer." Dennett now takes it that he has changed the task of relating two levels of explanation to answering "the question of how each bit of the talk about pains is related to neural impulses or talk about neural impulses." Making sense of this change of task takes us to step two.

Dennett's move now is to present a sub-personal functional characterization (a flow chart) which will capture all that is relevant to explaining consciousness, for instance, our awareness of pain. As a model for his introspective investigations he sketches a machine with a print-out capacity that enables it to "report" its "mental experiences." Construing these reports as analogs of a person's introspective reports, Dennett coins a technical sense of awareness:

\[ \text{A is aware that } p \text{ at time } t \text{ if and only if } p \text{ is the content of the input state of A's 'speech center' at time } t. \]

Use of this notion of awareness, he alleges, enables us to bridge the gap between the personal and the sub-personal levels of explanation. Since the speech center is construed functionally as a black box in a flow chart, we are to explain a person's awareness via a sub-personal account in
which we treat the content of his introspective reports simply as the output of the speech center. What the input to the speech center is, is that to which we have privileged access and which the pro-qualia intuition renders as qualia—+pain+. Dennett's radically counterintuitive (positive) position now surfaces:

The view that I wish to defend is that our privileged access extends to no images, sensations, impressions; raw feels, or phenomenal properties, but only to what might be called propositional episodes, instances of one's being aware [aware] that p . . . I call them judgments . . . I am left defending the view that such judgments exhaust our consciousness, that our individual streams of consciousness consist of nothing but such propositional episodes. . . .26

When we are in pain, then, Dennett's view is that we simply find ourselves wanting to say certain things, while engaging in other sorts of behavior ( . . . the baby with the bath water . . . ). But, and here there is a link with C in step one, the +pain+ itself is a posit of our folk psychological theory (which is not explanatory) and an unintelligible extra something at the sub-personal level. Thus, Dennett suggests we replace our personal level account by a sub-personal one in which there are no +pains+. What troubles him most about our homespun intuitions concerning qualia, especially +pain+ (here arises a variation on step one), is that they are fraught with inconsistency so that,
in fact, there is no true theory of pain. He elaborates:

\[\ldots\text{we find ourselves wanting to say}\]
\[\text{all these things about what is going on}\]
\[\text{in us; this gives rise to theories we}\]
\[\text{hold about how we come to be able to do}\]
\[\text{this—} \text{for instance, the notorious but}\]
\[\text{homespun theory that we "perceive" these}\]
\[\text{goings on with our "inner eye" and that}\]
\[\text{this perception grounds and explains the}\]
\[\text{semantic intentions we have}.^{27}\]

Still, however, it is a difficult leap from talk of how we do this to the elimination of that of which we are aware. Lots of what I want to say (as well as what I do) can be explained as being due to my awareness of a particular +pain+. As I understand his reasoning, there can be no true theory of pain since an adequate theory must preserve our entrenched intuitions (that which we care about), and in the case of pain we have equally strong but inconsistent intuitions. (The point about entrenched intuitions is a recurring theme; cf. Nagel's "What is it Like to be a Bat?" and Sellars' "Science, Sense Impressions, and Sensa: A reply to Cornman," and chapter 5 below. \(^{29}\)) Using the example of subjects given drugs to relieve pain (+pain+), Dennett argues that ordinary ways of talking (and thinking) allow us "to speak of drugs that render one insensitive to the pains that may persist." \(^{30}\) Such subjects, reports Dennett, often claim to be in pain but not to mind it, not to find it awful (not to notice the hurtfulness?). Why, however, are we to take the
introspective reports of subjects given drugs at face value?

My intuitions, gleaned by trying to conceive of myself as subject, are that they are not aware of +pain+, but are aware of other attendant features—perhaps they are aware of a throbbing, as when aware of one's pulse, that continues but no longer hurts (is no longer hurtful). That is to say, why not consider them as interpreting their inner state (the throbbing) as being just as it was except there is no +pain+, but reporting (drugs cause bizarre behavior) that the +pain+ does not "bother" them at all? Not having been a subject, my intuitions still lead me to say of them (the subjects) that their inner state does not "bother" them because there is no +pain+. Perhaps after having been a subject, I would tout a slightly different line (having myself been given drugs); in any event, however, the proper conclusion would seem only to warrant the claim that privileged access does not extend to drugged individuals (their reports lose veracity). At any rate, it is not at all clear that it is "equally ordinary" to speak of pains that do not "bother" one at all. Hence, I do not see that Dennett has established that "ordinary usage provides [equal] support both for the view that for pains, esse est percipi, and for the view that pains can occur unperceived."31
Nonetheless, Dennett suggests we abandon one set of intuitions in favor of another—his theory's preferred set. With regard to any flow chart—pain program—we are told that "[t]he only constraint on our design will be that it accommodate the known and presumed phenomena."\(^{32}\) We now see that he is speaking literally when he views the new task as that of explaining talk about pain. Indeed, as he views it, the phenomenon to be explained is pain talk, be it our own or another's; no longer is the data to be reckoned with the feel of pain (+pain+). Rather, it is that to which we have direct access: in our own case, the input to our speech center, our judgments, in the case of an other, her behavior, including her verbal reports. In this way Dennett hopes also to account for the asymmetry encountered earlier, for we are to some extent in a privileged (though not incorrigible) position with regard to the contents of our awareness, namely, what we want to say.\(^{33}\) Has something been left out? Consider:

I must grant, though, that at first sight my theory's domain of direct, immediate consciousness seems catastrophically underpopulated: there are no colors, images, sounds, gestalts, mental acts, feeling tones or other Proustian objets trouvés to delight the inner eye; only featureless— even wordless—conditional-intentions-to-say-that-p for us to be intimately acquainted with.\(^{34}\)
Being aware that such an account appears to leave something out—what it is like for us to, say, feel pain—Dennett recommends that we abandon our intuitions on this point, since they are inconsistent anyway, and opt for his theory and its information-flow model of awareness. But, unless step one is plausible, this suggestion does not do justice to the phenomena (+pain+) we, as introspectors, all experience. As I shall argue that step falters, I find solace in another functionalist's response to Dennett's way with qualia. William Lycan writes:

What bothers me more about Dennett's account is that it deals only with awareness qua propositional attitude, and does not even address the kind of difficulty that I think Block is trying to bring out. The feel of pain, the sweetness of a taste, and the grainlessness of phenomenal color are not, or at least not prima facie, propositional at all.35

Justifying this accusation amounts to giving credence to the strength of the pro-qualia intuition and depends on the ability to show that step one fails. It is to this task that we at least turn.

As I summarize Dennett's step one, I shall argue that his theoretical concern is too limited in that it centers on answering one question—how do we distinguish +pains+—at the expense of another equally legitimate question—what are +pains+? His attempt to dissolve the latter question by
relegating it to the personal stance (category) amounts to an ignoring of part of the data base for a satisfactory theory of mind. With regard to intentional properties, he argues that the personal stance is not explanatory because it must assume what it is charged with explaining, namely, intelligence. This move is not available qua phenomenal properties since our concept of pain includes +pain+, and unlike that of intelligence, finds its home in a different point of view. By ignoring the asymmetry between the introspective ("oneself") point of view and the scientific, objective ("the other") point of view, he succeeds only in slighting the data of the introspective viewpoint, thus making the move from the Manifest to the Scientific Image a precipitous leap in the dark.

As we see from P₁, Dennett maintains that to do any work in a theory, +pain+ must be identified. An explanatory theory, on my understanding, is charged with the task of answering questions; Dennett's question is: how does a person distinguish a +pain+, say, from a +tickle+? To be identified, +pain+ must be described or ostended (P₂). To describe, for Dennett, is to give identity conditions—to state the criterion applied in the act of distinguishing. There are, however, no identity conditions for +pain+ since a person does not distinguish it by appealing to a criterion; there is just the brute capacity to so discriminate. A bit
of caution is needed here, for certainly I do have a criterion for when to use "pain", viz., when +pain+ is part of the content of my consciousness. I can have a doubt as to whether something I am experiencing is a pain or not only because I have forgotten the meaning of "pain", without this entailing the bizarre claim that I lack an idea of what a pain is like. Moreover, should I need to decide whether or not a sensation is of pain, I can wound myself and experience a typical case of pain. I am able to distinguish a +pain+ from a +tickle+; what I can't do, and this is Dennett's point, is state in words how I do this—I can't (no one can) give directions that will enable another (machine?) to do this. Dennett's conclusion is that +pain+ can only be defined circularly, for when saying how a +pain+ is distinguished from a +tickle+ one must refer to an indefinable characteristic, namely, +pain+. Thus, the "mechanical question, how is it done? is blocked."

With regard to ostensive "definition", though it may enable us to come to properly use a term, Dennett maintains it will not serve the purpose of identification, for we cannot distinguish the quality (+pain+) from the discriminating of it. That is, we cannot "point" to a quality of the +pain+ which will allow us (to tell how) to pick out +pain+; we can only ostend the +pain+ as discriminated (cf. the case of the congenital non-pain feeler). Therefore,
since we can't identify (tell how to pick out) +pain+, it cannot do any work in a theory. But note that, as I have structured Dennett's argument, this only serves to show we are unable, from the personal stance, to answer the mechanical question.

To see Dennett's motivation for banishing +pain+ from the sub-personal level answer to the mechanical question, we need only glance at step two: unless we can write a program (construct a stepwise set of directions) that will account for the how of +pain+ discrimination, we will not have a functional characterization to be realized and thus, no functional explanation of our capacities. On the other hand, if we include +pain+ in the sub-personal account, the mechanical answer will be sullied by emergent non-physical properties. By eliding +pain+, Dennett seeks to resolve the ontological/causal issue as well as bridge Gunderson's distinction between program receptive and program resistant features of mentality.40 Motivation, however, is not argument; though it would make matters "easier" if there were no +pains+ to account for, his argument that there are none is an instance of an asymmetrical gaffe—a failure to recognize robust data.

Consider my (you consider your) own case. If a +pain+ is to be identified by me, then the fact that I have the capacity to recognize (again and again) different occurrences
of +pain+ (to identify various feels as +pain+), indicates that I (you, most of us) can ostend the pain quale. An exception would be the person congenitally unable to feel +pain+ (cf. Miss C.); no description would enable her to pick out (her own) +pain+, she has none, nor to know (by acquaintance) what +pain+ is, what "pain" refers to. Only ostensive "definition" allows one to know what (the) pain (data) is; ostensive "definition" appears nearly universal since most of us know what (the) pain (data) is.

It is this data that raises the question I take to be the most difficult for a theory of mind: what is +pain+ (the existence of which we are almost all aware)? How absolutely implausible it is to be told that the only difference between myself and one congenitally unable to feel pain, is that I have and she has not the inclination to say certain things and engage in certain patterns of behavior not accompanied by +pain+, but by certain abstract functional properties. Only by ignoring the introspective point of view, a point of view I cannot ignore, is this remotely plausible. If Dennett did not say that he means what he said in step on (see opening quote, p. 95 above), I would be loath to draw the conclusion that he seriously maintains there are no +pains+; but he does.

My analysis of his argument indicates that Dennett takes the theoretical question of concern to be the
mechanical question; obviously one way to avoid the earlier ontological/causal quandary is to ride roughshod over serious puzzles concerning the nature and status of sensory consciousness by, in effect, ignoring the asymmetrical distinction raised by the data of introspection. That Dennett does the latter is seen by the fact that, for him, to identify is to give identity conditions rather than to have the capacity for recognition (again and again). Focusing concern on the data of introspection, we see that I, and you, can identify +pain+, though none of us is able to say how. True, the mechanical question wants theoretical answer, but to dissolve our ontological/causal worries by arguing for (what amounts to) the claim that the question, "what is +pain+?", flows from a category mistake, is not plausible (if intelligible). Dennett, of course, would not be impressed by such complaints; part of his ploy is to admit that his theory is counterintuitive—that, however, does not make it any more plausible. Can Dennett nonetheless comfortably maintain his line? I think not, as a further argument indicates.

**Posits: For the Purpose of**

I turn now to a further criticism of Dennett, adapted from George Pappas. For Dennett, positing a quale which does no work in a theory "is thus insisting on an
unintelligible extra something." As I understand him, *pains* are like "mental images themselves . . . the creatures of a "posit", an inference or extrapolation exactly analogous to Hume's "posit" about external bodies." They are, though, posits that are not necessary; the positing of them multiplies entities beyond necessity. So, the question to be dealt with is, why aren't they necessary—necessary for what? and how are we justified in eliminating them tout court? (As I have argued above, they remain as data to be explained even if we have trouble understanding them as explainers in our mechanical/causal theory.) The proffered answer is to be unearthed from Wittgenstein's earlier aphorisms via Quine's brief for physicalism:

If there is a case for mental events and mental states, it must be just that the positing of them, like the positing of molecules, has some indirect systematic efficacy in the development of theory. But if a certain organization of theory is achieved by thus positing distinctive mental states and events behind physical behavior, surely as much organization could be achieved by positing merely certain correlative physiological states and events instead. . . . The bodily states exist anyway; why add the others?45

Dennett, of course, would distinguish himself from Quine by positing additional inner functional states; but his overall strategy is the same. Like Churchland and Churchland, Dennett construes our commonsense conceptual
framework as falling within Quine's broad sense of "science", thus that framework involves explanatory theory construction. Hence, our introspective reports are considered theory-laden to the extent that our sensations are posits. "Pain" does not refer, argues Dennett, because positing pain (+pain+) does not promote "our" explanatory purposes. (I assume here that our concept of pain, that which we feel (+pain+), is not obviously contradictory or inconsistent as his second version of step one claims.) To posit the existence of the pain quale, then, is to assume that there is some such thing for the purpose of providing an explanation of some data. Here, as I have argued above, is where the problems begin. Just what is the data to be explained? As we have seen from step two, Dennett maintains it is our and other's behavior, including our verbal behavior, namely, our introspective reports of that of which we are aware. Yet, to make this all the phenonema in need of explanation—all that there is—involves arguing for the principle:

Q Everything that exists is such that a person is justified in believing that such an entity exists only if the assumption that it does serves some genuine explanatory need.

Apparently Dennett just does not recognize (theory-blindness?) the feel of a splitting headache as data in need of explanation—he wants data to be statable facts; thus
+pain+ does not serve a genuine explanatory need. One may think this request that data be statable, a reasonable one. Thus, to maintain that X is in pain if, and only if X realizes (by virtue of N₂ realizing) functional state Sₚ, is to produce a non-accidental generalization that entails the statement that X is in the pain state. Yet the datum to be explained is not simply that X is in the pain state (or +pain+ has occurred), but rather that +pain+ is ____. That is, it is not only the question, "Why (causal) is X in pain?", that needs answering but also "What is +pain+?". Answering the former only tells us what is causally necessary for +pain+ to occur (emerge), not what +pain+ is. As I have argued, the inner causal happenings science describes are not constitutive of felt pain. For Dennett, the question, "What is +pain+?", is not legitimate because +pain+ is not in the data base. What, then, am I aware of when I withdraw quickly from contact with red-hot objects and continue to avoid them whenever it is in my power, if not the +pain+ that accompanies my pain state? Working backwards for a moment, let us realize just what Dennett is up to.

Should he be correct, I did not suffer from the +pain+ of a sore throat and headache this morning, though I did behave in certain cranky ways, took aspirin, cursed my fate and bemoaned my wanton ways. But I could not suffer from the feel of +pain+ because there are no +pains+, only my
behavior and my inclinations to say things. Dennett just denies that suffering involves *pain*, a view suspiciously like that of some Cartesians who denied feelings to animals, as Fontaine relates:

There was hardly a *solitaire* who didn't talk of automata. . . . They administered beatings to dogs with perfect indifference, and made fun of those who pitied the creatures as if they had felt pain. They said that the animals were clocks; that the cries they emitted when struck, were only the noise of a little spring which had been touched, but that the whole body was without feeling. They nailed poor animals up on boards by their four paws to vivisect them and see the circulation of the blood which was a great subject of conversation.49

Descartes, in commenting on his view that the brutes do not have feelings says, "[t]hus my opinion is not so much cruel to animals as indulgent to men . . . since it absolves them from suspicion of crime when they eat or kill animals."50 Obviously, Dennett wishes to avoid such company. Compare:

If and when a good physiological sub-personal theory of pain is developed, a robot could in principle be constructed to instantiate it. Such advances in science would probably bring in their train wide-scale changes in what we found intuitive about pain, so that the charge that our robot only suffered what we artificially called pain would lose its persuasiveness. In the meantime (if there were a cultural lag) thoughtful people would refrain from kicking such a robot.51
It is not clear, however, how successful he can be at avoiding the company of such Cartesians, for as long as he maintains there are no +pains+ and suffering does not involve +pain+, his view seems tainted.

Nevertheless, my awareness and suffering of my early morning +pain+ gives me pause here, and make the following quite plausible:

What we are aware of is something such that we are justified in believing it to exist, and no sound scientific reasons are needed to justify this belief.52

So, unless Q is true, my justification for holding that there are +pains+ does not depend upon +pain+-positing discharging an explanatory task. Even if the explanatory advantages gained by such an assumption (viz., the reason why I avoid red-hot objects) were lost, I would still be justified in believing there are +pains+ because I feel and suffer them.53 Hence, unless Q is true, the pro-qualia intuition remains intact. But Q has serious problems.

Consider any existential statement. If Q is true, then any such statement we accept is justified on explanatory grounds. But surely

[w]e cannot seriously hold that every sentence we accept is justified on scientific (explanatory) grounds, on pain of admitting either the circularity of all justification or an infinite regress of
justification. There must be some data sentences that [we] (justifiably) accept simply because they seem to express truths that are hard to ignore. What guarantees that none of these data sentences will be an existential sentence?—in particular, a sentence asserting the existence of something that we are aware of?\textsuperscript{54}

My earlier argument against step one shows that there is such data and thus, such data sentences. Q, however, restricts us to the contention that a belief, if it is to be justified, must serve to explain other beliefs. Hence, we are restricted to the claim that a statement is justified (for a person) only if that statement serves to explain, or helps to explain, some other statements.\textsuperscript{55} Pappas nicely summarizes the infinite regress difficulty which troubles Q:

Thus, let $p$ be some putatively justified statement, and let $q$ be some statement which $p$ explains or helps to explain. Either $q$ is itself justified or it is not. If it is not justified, it is hard to see how any justification will accrue to $p$ in virtue of $p$'s explaining $q$. On the other hand, if $q$ is justified, it is so not only if it explains or helps to explain some other statement $r$. The same considerations which affected $q$ may be brought forward regarding $r$, so that, it would seem, we are led to yet another statement, $s$. And so on.\textsuperscript{56}

Without Q, however, the Quinean gambit appears to fail and the pro-qualia intuition survives. But, as Pappas points out, the infinite regress difficulty wielded against Q may itself be inconclusive.\textsuperscript{57} For one may respond that
if it's the case that if $p$ is to be justified for me, then $p$ must explain $q$, then $q$ need only be initially credible for me (not itself justified). Thus, $q$ would not itself have to be an explainer and the infinite regress would be halted. Perhaps, though I'm not sure what to say about the strength of the justification that would accrue to $p$ in virtue of explaining a $q$ that is (only) initially credible. Since the latter notion appears to involve us in plausibility considerations, it may well turn out that the strength of the justification that accrues to $p$ will be a function of the degree of plausibility of $q$. And this will lead us to a case by case consideration of our $p$'s and $q$'s.

In the present case, let "$q$" = "my screams, the withdrawal of my hand from the red-hot object and my ensuing pain-behavior are all accompanied by the +pain+ I feel" and let "$p$" = "there are +pains+ which people dislike and which typically accompany screams, withdrawal behavior, etc., and which are themselves caused by touching red-hot objects." Given the arguments above, $q$ is certainly initially credible; since I take it that $p$ explains or helps explain $q$, $p$ is justified for me (and, I assume, you).

So, if +pain+ is not a posit qua $Q$, and even if +pain+ did not figure in explanations of human behavior (itself a dubious assumption), our awareness of +pain+ would leave as unexplained data a qualitative residue. It is that residue
that accounts for the fact that people in pain (+pain+) are sufferers. Should one support Q via the initial credibility line, then explanations in which +pain+ figures would be justified—some of these, such as p—would commit us to the existence of what we know so well, +pain+. In the latter case, we would still have our ontological/causal worries. In either case, the mind-body problem remains and the inverted qualia objection is thus unscathed (so far). 58
4. PUTTING QUALIA ON THE LEVEL

Block's challenge remains:

If inverted qualia are possible, then mental state 0 is not identical with any functional state, and if functionalism claims that every mental state is a functional state, functionalism is false.¹

**Homunctionalism**

Lycan, though apparently more sensitive to the force of the pro-qualia intuition, nonetheless takes his cue from Dennett's homuncular theory of psychological activity.² Unlike Dennett, however, he is not out to deny that there are qualia,³ but to "defend an ontology of the mental which is functionalist in a more robust sense of the term function..."⁴ By identifying mental entities with items teleologically characterized via the roles they play in furthering the interests of a containing system (see below), Lycan hopes to appease the pro-qualia intuition and maintain the relationalness of functionalist explications. His tack is to recognize the reality of qualia at a certain level of organization. To obtain a clearer understanding of the homunctionalist (the term is Lycan's) strategy and the above pithy statement of the task, consider Dennett's proposal
that we view the person on the metaphor of a bureaucratic corporation:

A flow chart is typically the organizational chart of a committee of homunculi (investigators, librarians, accountants, executives); each box specifies a homunculus by prescribing a function without saying how it is to be accomplished (one says, in effect, put a little man in there to do the job). If we then look closer at the individual boxes we see that the function of each is accomplished by subdividing it via another flow chart into smaller, more stupid homunculi. Eventually this nesting of boxes within boxes lands you with homunculi so stupid (all they have to do is remember whether to say yes or no when asked) that they can be, as one says, "replaced by a machine." One discharges fancy homunculi from one's scheme by organizing armies of such idiots to do the work.⁵

Embedded in this tralatitious statement is the notion of a systematic explanation⁶ of the abilities of a person by appeal to the organization and abilities of the person's proper parts. Unique to this style of explanation is the notion that these parts interact (the homunculi "communicate" with one another) to serve the ends of the whole. So, at the "higher" level of organization, the whole person is treated as the containing system; its interdependent parts (the homunculi) interact to further the purpose of the person. Analogously, the workers on an assembly line work together, according to an organizational plan, to produce the finished goods thereby serving the ends of a corporation. In an
organism/system as complex as a human being the homunculi (functional components) are often organized into smaller systems, thus engendering a systematic hierarchy or multi-level organization of nested systems (boxes in a flow chart). Explanation proceeds, at the next lower level of organization, by reiterating the homuncionalist strategy until, as Dennett puts it, "a mechanistic view of the proceedings becomes workable and comprehensible." Interestingly, however, this top down explanatory progression is not reduction, that is to say, reduction in the sense that the lower level explanations supplant the higher level ones.

As Putnam argues, explanation is not transitive since certain systems can have behaviors to which their microstructure is largely irrelevant. Putnam's point is that "from the fact that the behavior of a system can be deduced from its description as a system of elementary particles, it does not follow that it can be explained from that description." As an example (Putnam's), suppose we have a peg and a board and want to explain

D: The peg goes through the square hole, but does not go through the round hole.

Consider the explanation

E: The peg and the board are approximately rigid; the peg goes through the hole large enough but not through the hole that is too small.
Here talk of microstructure is irrelevant. Consider explanation:

\( E^o \): Here we describe the board and the peg as clouds of particles such that the board = cloud B, the peg = cloud A, the round hole = region 1, and the square hole = region 2. Through an "heroic feat of calculation" on the position and velocity of each particle, we prove that cloud A will pass through 2 but not 1.

Given pragmatic constraint \( C \): "The relevant features of a situation should be brought out by an explanation and not buried in a mass of irrelevant information," Putnam concludes that \( E \) does but \( E^o \) does not explain why cloud A (the peg) passes through 2 (the square hole) but not 1 (the round hole). (Strictly speaking, this is incorrect; \( E^o \) does not explain \( D \) for us (given our current abilities), but to a Laplacean scientist of the future, \( E^o \) would serve as an explanation. We may still, of course, say that \( E^o \) does not (now) explain \( D \), but it is not so clear that, that \( D \) is not explained by \( E^o \) shows that explanation is not transitive, for \( D \) may well be explanable by \( E^o \) (in the future).)

In this sense, then, the higher level explanation is autonomous. Hence, complete reductions—where explanations couched in terms of theoretical physics supplant (translate via bridge rules into) all other explanations—is not a desideratum according to this strategy. (I'm not sure, however, that this rules out reduction to neurophysiology.)
Teleology (An Aside)

Though Lycan is himself a protagonist of the above strategy, he strays (wisely to my mind) from Dennett's instrumentalist treatment of qualia as mere *abstracta* to pursue a reductive type-identification of beliefs and other mental items, including perceptual and other "qualitative" states, with homuncurnal states. Putting my cards on the table, I suggest we identify a mental state-type with the property of having such-and-such an institutionally characterized state of affairs obtaining in one (or more) of one's appropriate homuncurnal "departments" or subagencies. . . . To be in pain of type T, we might say, is for one's sub-subpersonal Ø-er to be in characteristic state $S_T(Ø)$, or for a characteristic activity $A_T(Ø)$ to be going on in one's Ø-er.11

As Lycan recognizes, whether these are winning cards depends on how these teleologically characterized workings of the organism are handled. For example, what is the status of the claim that to be in pain (to be aware of +pain+) is to have homunculi pulling levels and throwing switches for the purpose of initiating appropriate behavior? At no point do such characterizations stop being teleological, period, and become mechanical, period. Rather, there is "a finely grained continuum connecting the abstract and highly teleological to the grittily concrete and only barely teleological."12 The idea here is that there is no objective functional/physical distinction. The link between levels is,
of course, explanatory and since lower level explanations (themselves teleological) do not supplant higher level explanations, there will be ineliminable reference, at all levels of characterization, to teleologically characterized workings of the organism. Though some levels will be only "barely teleological," it is not immediately clear how much of our understanding of successively lower and lower levels depends on "carrying along" the higher level teleological characterizations. That is, it is not immediately clear where talk of purpose becomes harmless. As Charles Taylor points out, the question of the elimination of the teleological is intimately bound up with the question of the reducibility of the intentional. Typically, the problem posed is that teleological/intentional characterizations cannot be interpreted in causal terms. Taking this problem seriously, one cannot help but view it as, at least, a prima facie difficulty for the notion that there is ineliminable reference, at all levels of characterization, to the teleologically characterized workings of the organism. Thus, Joseph Margolis complains:

The critical fact about Dennett's story is simply that, however attenuated, the stupid homunculi are ascribed intentionally qualified abilities. They process information. It hardly matters that, as their operations become more stupid (or routinized), they may be replaced by machines; for the informational content of what they do—and of what
the machine does that replaces them—is such only on the interpretation of some molar person to whom the relevant intelligence and understanding may be directly ascribed.  

Dennett offers no argument to show that, at the homuncular level, if molar human agents are to be analyzed without remainder, ascriptions of real psychological states, conscious and cognitively qualified, can be avoided. This is what we cannot be sure of in Dennett's introduction of idiot homunculi. If they are merely stupid, his reductive program fails, because their behavior will be explained by means of the same psychological model (normally reserved for molar agents) that he wishes to eliminate. . . . and if they are less than stupid, utterly lacking in intentionally qualified states, then his reductive program will fail again, because neither he nor we know of any way to reduce the intentional to the non-intentional.

Margolis' complaint seems to be that the task the homunctionalist has set for himself requires prodigious bootstrapping, since homunctionalists must use the very same intentionally charged characterizations that they wish to explain in order to explain. This is, of course, an analogue of a similar problem facing the neuroscientist in his quest to understand the functioning brain, namely, the scientist must use the same structures and cells he wishes to comprehend in order to comprehend. Although the homunctionalist problem seems more severe, it is not clear just how severe it actually is. It is certainly a prima facie difficulty and one recognized by Lycan and Dennett. Both are inclined
to believe that evolutionary considerations can solve this problem. Lycan writes:

On this general issue [teleology] I have little of my own to contribute. I hope, and am inclined to believe, that the teleological characterizations that homunculicism requires can be independently explained in evolutionary terms.17

Dennett sets one wondering with his view that:

The final reductive task would be to show not how the terms of intentional system theory are eliminable in favour of psychological terms via sub-personal cognitive psychology, but almost the reverse: to show how a system described in psychological terms could warrant an interpretation as a realized intentional system.18

It strikes me that there is an explanationist gambit afoot with regard to the notion of reduction on which Lycan and Dennett both lean. Wimsatt19 attempts to drive a wedge between the ontological and explanatory dimensions of reduction, where the explanatory dimension is considered more important. Adequate examination of the possible success of this program, however, calls for a more detailed scrutiny than I am able to give here. Putting this problem on hold, let us return to Lycan’s way.

Pain as Analysable

For Lycan’s version of materialism (assuming a successful resolution of the teleological issue), such irreducibility
is not seen as a difficulty "so long as our system of institutional categories, and our system of physiological categories, and our system of physical categories are just alternative groupings of the same tokens."\(^{20}\) Since Lycan assumes that institutional tokens "are reducible in the sense of strict identity, all the way down to the subatomic level,"\(^{21}\) he sees double aspect worries as not being raised. Now, if I read Lycan correctly, it is here that we hit upon the locus of his departure from Dennett's instrumentalism. As we have seen, Dennett claims that "different categories are no better than different Cartesian substances unless they are construed as different ontological categories and only one category of terms is referential."\(^{22}\) Lycan seems to maintain, however, that "pain" will enter into our explanations at various levels of organization depending on whether our concern is with the more behavioral or more feely aspects of our concept of pain. That is, it may be that we'll classify an internal state of an organism "as a pain" based on behavioral data (e.g., withdrawal behavior), but "distinguish the feels of pains according to the states' physiological bases." Yet merely distinguishing the feels of an other's pains (+pains+) according to physiology is perfectly compatible with interactionism or epiphenomenalism—unless a robust reduction is in the offing. As has been suggested (and will be argued in Chapter 6), we distinguish
our own pain states, not according to physiology or function (or behavior), but via what it is like to experience such states. Thus, according to our concept of pain, "pain" refers not to behavioral items, physiological items or functional items but to felt items picked out ostensively. Lycan concedes that it would "follow that a state that feels like a pain state of mine might in a differently organized creature be a mental state of some kind other than pain; some philosophers may find this crassly counterintuitive."24 Indeed, I do, especially if we stray from science fiction and limit our domain to organisms in an evolutionary milieu.

Alternatively, then, Lycan suggests it may be "that the distinctive quale associated with a feeling of a certain type is really the coincidence or superimposition of a number of distinct, individually manageable homunrectional features."25 But the claim that these distinct homunrectional features are individually manageable amounts, it seems, to the claim that qualia are not unanalysable simples. Lycan, suggests Keith Campbell,26 attempts to analyze qualia as having propositional character—despite his protestation against Dennett (p.109 above), Lycan relies on Dennett's pain data and maintains:

What the drugs seem to be doing is splitting off components of the subjects' phenomenal experience of the pain, by splitting off component subsubroutines of its rather
complicated functional basis. And if this is so, it follows that our phenomenal experience of pain has components—it is a complex, consisting (perhaps) of urges, desires, impulses, and beliefs, probably occurring at quite different levels of institutional abstraction.\(^{27}\)

For Lycan, "this makes it reasonable to suppose that some proper . . . multi-leveled subsequence of the relevant complex of functional goings-on is both necessary and sufficient for the occurrence of the pain. . . ."\(^{28}\) Again, where "pain" refers to (+pain+) and not mere behavior, etc., functional goings-on being necessary and sufficient for the occurrence of pain (+pain+) is compatible with interactionism or epiphenomenalism. To see this, compare how Lycan turns the tables on Block's\(^{29}\) cases of merely functionally complex organizations in which qualia are alleged absent despite meeting functional specifications.

"A neuron is just a simple little piece of insensate stuff that does nothing but let electrical current pass through it from one point in space to another; by merely stuffing an empty brainpan with neurons, you couldn't produce qualia--immediate phenomenal feels!"--But I could and would produce feels, if I knew how to string the neurons together in the right way. . . .\(^{30}\)

So, if such neurons realized the pain-profile for an organism, then, even though each neuron was devoid of qualia, appropriately connected up they would give rise to ("produce") qualia--immediate phenomenal feels. Though these qualia may
be said to supervene (see below) on the neuroanatomical, they still lack explanation (unless reduced). Given the earlier quote, it seems that, for Lycan, our phenomenal experience of pain "is a complex, consisting (perhaps) of urges, desires, impulses, and beliefs. . . ." We are now back to the claim that any system with the appropriate complex of urges, desires, impulses and beliefs will be a pain feeler. Keith Campbell maintains these items themselves must be interpreted bloodlessly:

The impulses, desires, etc., must not be given their own qualia of felt urgency, longing, enduring, or wishing for relief, else the treatment will not eliminate qualia. Each of the component impulses, etc., must on its own be entirely bloodless. But now the proposal's plausibility evaporates. It makes being in pain just like being in a difficult and unwanted position in chess.31

I think Campbell on the right track here, for some sort of elimination does appear in the offing. If Lycan maintains that a qualitative state is but an assemblage of individually bloodless items, then he is simply relying on the logical principle that what is true of the whole need not be true of its parts. Still, however, this runs close to begging the question for it is prima facie plausible that if the mere organized collection of cells cannot feel pain, insofar as a person can feel pain, he is more than his cells and their organization. And where the antecedent of this latter
conditional is denied—as appears in the case of the brainpan and its neurons—+pain+ (quality-object) arises as an emergent phenomenon. Perhaps this helps explain the move (back) to the propositional attitudes—in the hope of allowing "qualitative" features to be expressed in terms of information transformation. But this move seems implausible; it seems but an attempt to eliminate data.

Suppose that the propositional attitudes do inherit qualia—urgent desires do have feels. I find this mysterious; is the +pain+ I feel now, upon touching a red-hot iron, to be analysed as an urgent desire that I remove my hand? If so, it is successful as an analysis only if "urgent desire" is taken to have as referent +urgent desire+—otherwise we are right to wonder where the qualia have gone. It seems plausible, then, that if propositional attitudes are imbued with qualia, the what it is like feature of having an urgent desire still remains. The congenital non-pain feeler would lack the urgent desire to remove his hand from the red-hot iron and despite total physical/functional knowledge would not know what it is like to have such a desire. That it feels like something to touch a red-hot iron cannot be doubted. If the urgent desire is to have a feel in the same sense, then what it is like will not be known by the non-pain feeler. If qualia are to be reduced, it does not appear that shifting them to the propositional
attitudes will help. And if the claim is that "pain" at t just is a complex consisting of propositional items (sans the what it is like feature), then, as Campbell suggests, the proposal's plausibility evaporates—being in pain is not like being in a difficult and unwanted position in chess unless the latter itself has qualitative features just as much in need of explanation.

We can now tie into the teleological/intentional problem raised above. On whatever level we focus "pain" refers to a certain teleologically characterized homunfunctional state. Certainly this is, at first blush, quite counter-intuitive with regard to the feely aspect, "pain", unless we understand the homunfunctional state as giving rise to, or itself containing, the qualitative feature of concern. Thus the question remains, what is the nature of that to which "pain" ("+pain+") refers? The teleological/intentional difficulty, even if solved, appears to leave a happy answer to this question out of reach. Consider Dennett:

The principles of evolution proposed to explain learning and discrimination in the brain have the capacity to produce structures that have not only a cause but also a reason for being. That is, we can say of a particular structure that the animal has it because it helps in certain specified ways to maintain the animal's existence.34

On the assumption that teleological explanations are eliminable in favor of non-teleological explanations, a
difficulty still remains for the causal story (ignoring the conundrum that the reason for being of certain structures is put into the explanation by us, the molar agents to be explained). Where "in order that" captures the notion of "reason for being" (or "for the sake of"), consider the following schema of explanation:

\[
\begin{align*}
\text{molecular} & \quad \text{in order that} \quad \text{structure} \quad \text{in order that} \\
\text{nerve assemblages} & \quad \text{in order that} \quad \text{+pain+} \\
\text{withdrawal} & \quad \text{in order that} \quad \text{behavior} \\
\text{avoidance} & \quad \text{of damage} \quad \text{in order that} \quad \text{...} \\
\text{continuation} & \quad \text{in order that} \quad \text{of species}
\end{align*}
\]

Notice that the "in order that" locution must (assuming reduction) have a causal reading if we are to make sense of the role of +pain+ in the evolutionary story (else +pain+ is an epiphenomenon). Unless we reduce a particular occasion of +pain+ to, in Sellars' terminology, physical characteristics, it begins to look as if we are formulating a mentalist theory of the brain, not a materialist theory of the mind. Where +pain+ has a causal role it would follow that there can be states of a person which are not complex motions of, say, (bloodless) atoms; sensory items (construed as or as having phenomenal properties) are then allowed to
play an essential causal role in the behavior of the bodies of sentient beings. Unless, that is, such items can be reduced or eliminated.

Supervenience

We must know, therefore, that although the mind of man informs the whole body, it yet has its principal seat in the brain, and it is there that it not only understands and imagines, but also perceives; and this by means of the nerves which are extended like filaments from the brain to all the other members, with which they are so connected that we can hardly touch any part of the human body without causing the extremities of some of the nerves spread over it to be moved; and this motion passes to the other extremities of those nerves which are collected in the brain round the seat of the soul. . . . But the movements which are thus excited in the brain by the nerves, affect in diverse ways the soul or mind, which is intimately connected with the brain, according to the diversity of the motions themselves. And the diverse affections of our mind, or thoughts that immediately arise from these motions, are called perceptions of the senses, or, in common language, sensations.\(^\text{36}\)

Lycan maintains that the mental supervenes on the psycho-functional and concedes that Descartes held a similar view, where the mental supervenes on the neuroanatomical. But Descartes was a dualist and the items that supervened were "eerie" mental items which fostered ontological/causal difficulties (what is the nature of the "intimate connection" of the mind with the brain?). So a question arises: How is
Lycan's view dissimilar and how is my particular pain now, construed as a supervenient phenomenon to be rendered ontologically/causally comprehensible? One way to reckon the supervenience relation is hinted at by Donald Davidson, who maintains that his view (anomalous monism) is non-reductionist but.

... consistent with the view that mental characteristics are in some sense dependent, or supervenient, on physical characteristics. Such supervenience might be taken to mean that there cannot be two events alike in all physical respects but differing in some mental respect, or that an object cannot alter in some mental respect without altering in some physical respect.

The first thing to point out is that Lycan's view must be quite similar as far as its logic goes, for in order to maintain his materialist version of functionalism he must hold that the supervenience relation is transitive (see below). Just as there can be no mental occurrence unless a functional characterization is realized, there can be no realization unless the functional supervenes on a structure of enough complexity, organized in the appropriate (teleological/evolutionary) way to do the realizing. That is, if the mental supervenes on the functional, then since the functional supervenes (if we are to maintain materialism) on the physical (it supervenes on structure and the categorical base of the structure is physical), the mental must supervene
on the physical. Hence, for any particular instance of +pain+ in S at t to supervene on the functional, it must be that the functional syndrome (institutional type) identified with this sensational supervenes on the particular neuroanatomical items in S at t which realize or embody this functional syndrome (the tokens which play the appropriate roles must be physical tokens), otherwise we would have functionalism sans materialism. (Of course we should understand the particular neuroanatomical items that do the realizing under their physical guise so as not to exclude nonhuman systems from having mental states.) If the above is correct, then a particular mental occurrence supervenes on the physical by way of supervening on the functional which supervenes on the physical. So Lycan and Davidson are close cousins in this matter.

Let me quickly support the claim that if +P+ supervenes on "F", then since "F" supervenes on N, +P+ supervenes on N.

The first step is to show that the supervenience relation is transitive, the second is to argue that F supervenes on N. The quick way to accomplish the first step is to lean on Kim, who offers a definition of supervenience (see below) and then states: "... this definition more directly states that each (instantiated) property in the supervenient set [e.g., F] has a sufficient condition in the supervenient base set [e.g., N]." Given this, it
follows that $F + P$, $N + F$ and thus, $N + P$. So, if $F$ supervenes on $N$, $+P$ supervenes on $N$ (Lycan claims that $+P$ supervenes on $F$). This takes us to step two: is $N$ sufficient for $F$? Certainly if the tokens constituting $N$ are properly organized, we cannot fail to have function $F$. After all (using Lycan's example), if a collection of molecules is properly organized, we cannot fail to have a metal strip with an articulated flange which has the function of a key. As Campbell points out:

There is indeed a sense in which function is supervenient on structure. Given structure $S$ in circumstance $C$ one cannot fail to have function $F$. There is a necessity about camshafts in their proper place in engines working as valve lifters.40

Hence, $F$ supervenes on $N$ and $+P$ supervenes on $N$. Now I turn back to supervenience and necessity.

Davidson's claim, however, is problematic: how are we to understand the claim that an organism cannot differ in some mental respect without differing in some physical respect? It would not be wise to understand this as logical necessity, where that involves a priori entailment from mental to physical, for that would amount to an a priori denial of the possibility of disembodied minds which change mental set. Likewise, the claim that there cannot be two organisms alike in all physical respects but differing in some mental respect, cannot happily be understood as
involving logical necessity for this would force the view
that a complete physical description of Jones would entail,
a priori, that she was, say, in pain—a form of analytical
reductionism.

Nor, it seems, will nomological necessity suffice; not
however, simply because I can conceive of both the above
claims being false (which I can). For it to be nomologically
necessary that there is no mental difference without a physi­
cal difference (and vice versa), there would have to be laws
(contra Davison) supporting correlations (parallelism? dual
aspect?). Davidson's claim is that there are only physical
laws and that pain, say, as a causal factor must receive a
physical description. But as Kim notes:

It is difficult to see how such neuro­
physiological descriptions can be chosen
apart from our discovery of psychoneural
correlations between phenomenal mental
events, such as pains and tinglings, on
the one hand, and certain underlying neural
processes on the other.41

As Kim goes on to point out, such correlations only
serve to exacerbate the general problem of psychophysical
causation. Of course we need not commit to laws of the
psychoneural/psychophysical sort; as Lycan argues,42 the
road to token-identity may incorporate psycho-functional
laws. Thus the proposed type-identification of mental
states with functionally characterized states (of rendering
psychological characterizations as functional characterizations). Although functional types are irreducible to physical types (cf. problems generated by equipotentiality), any realization or embodiment of a particular homunfunctional state must be by (to retain materialist proclivities) physical items. While the functional type is not identified with its realization, a particular institutional token is "reducible in the sense of strict identity, all the way down to the subatomic level" (token-identity). If we lean on Dennett's analogy of the nesting of boxes within boxes, we see that Lycan's strategy is to begin with institutionally characterized states of affair and break them down into subagencies until we are far enough down to speak of the realization of the institutional type by a physical token.

So it begins to look, if we are to have materialism, as though the institutional tokens of the institutional type identified as the pain state at t must be purely physical, say a subset of the noiceptive neurons firing at t. To say these tokens are purely physical is to maintain, as does Lycan, that these tokens are reducible in the sense of strict identity all the way to the subatomic level. Thus, all the properties of these tokens must be such that each property instance is reducible to instances of purely physical micro-properties of the constituents of these tokens. Since pain is alleged to be a property of these tokens, each instance of
pain must be nothing but an instance of such microproperties. As we have seen, pain states exemplify hurtfulness and particular instances of felt pain (+pain+) are quality-objects of which we are aware when in the pain states, i.e., when (let us suppose) some subset of our noiceptive neurons is firing. So it must be maintained that +pains+ are nothing but property instances of such firings. Let us understand the above in summary form as the claim that the institutional tokens realizing an institutional type are reductively identified with the physical tokens (microproperty instances) constituting the physical state (subset of firing noiceptive neurons) which realizes the functionally characterized institutional type (pain). Lycan's claim, then, is something like the following:

L: For each psychological state M there is a functional state F such that, as a matter of law (necessarily?), a state of type M occurs in an organism (system) at a time just in case a state of type F is realized by the organism (system) at the same time. And the institutional tokens constituting F are reductively identified with the physical tokens constituting the underlying (physical) structure which realizes F.

With this in mind, let us see how Lycan fares with the above problems facing Davidson.

Lycan would maintain that (i) an organism cannot differ in some mental respect without differing in some functional respect. Understanding this to involve logical necessity
does not amount to an *a priori* denial of the possibility of disembodied minds, for it is logically possible that non-physical items will realize the functional syndrome—if the notion of an immaterial cause is coherent. It is not plausible, however, to maintain that a complete functional description of an organism in functional terms would *entail* the appropriate mental descriptions *a priori*. If this were the case, then the *logical* possibility that Sid and Amy are functionally equivalent yet Sid sees green what Amy sees red, would be ruled out *a priori*. But it is to be conceded that it is a logical possibility. Thus, the claim (ii) that there cannot be two organisms alike in all functional respects but differing in some mental respect, cannot be understood as involving logical necessity (at least not happily understood). Still this is not so telling an objection. It still seems implausible that one's spectrum could be inverted relative to the entire molecular constitution of the central nervous system—albeit logically possible. Again, construing the functionalist as materialist, as in L, brings Lycan and Davidson close together. So, what sense can be given to the claim that if Sid and Amy are functionally/physically identical (within the relevant parameters), their qualia *cannot* be inverted? Should the necessity be nomological, then we have but law-like correlations which are compatible with a number of non-materialist theses. Supervenience, to be of
use to the materialist and not merely label the difficulty facing the interpretation of such correlations, must have an adjunct notion, namely, reduction. Let us pursue the necessity involved.

It may be that there is a sense of necessity, namely, metaphysical necessity (à la Kripke) that can be invoked here. As we shall see, however, such necessity assumes reduction of the mental to the functional-cum-physical. Suppose we agree that that of which I am aware now (+pain+) is named "+_P+", then +_P+ would (following Kripke) be a rigid designator. In any possible world +_P+ will designate the what it is like feature of my pain state. For any mental difference, Lycan maintains, there will be an associated functional characterization _F_. Let us see what the nature of this association will have to be.

Obviously "_F_" does not rigidly designate particular physical items, for in one world it may pick out neural structure _N_ and in another world an entirely different structure, that had by any alien, say; that is, what realizes _F_ may vary from world to world. Of course if "_F_" designates a type, it does so rigidly. So where we are dealing with a type of pain state, _P_, _P_ will be associated with a functional characterization such that in any possible world in which there is _P_, its functional characterization must (if we are to have materialism) be realized by some physical
structure or other, say $N$, reductively identified with purely physical microproperty instances. Since Lycan claims that "mental entities [construed as types] have no physical essences, but they do have functional essences," it must be that $P$ is realized by physical property instances of token $N$, which realizes $F$, for Lycan claims the pain state, $P$, is identical to $F$ (to be in pain is to realize $F$; we are, it seems, to understand "being in" as "realizing"). So for particular pain states like $P$, their association with their functional characterizations (institutional types) is one of identity. What of the property instance $+P+$? What is the association between $+P+$, functional properties and physical properties (instances, of course)?

To be consistent it seems Lycan would maintain that any property instance of the property of being in pain must be reductively identified to functional property instances which are in turn realized by physical microproperty instances. It would appear that what is meant by "functional property instance" is a low-level functional syndrome gleaned by reiterating the homunctionalist strategy of "looking" deeper inside to lower and lower levels of organization— in short, a description of a low-level of organization. These low-level descriptions are to be construed as true of (realized by) microproperties of physical items.
To say there cannot be a mental difference without a functional difference, in the case of particular +pains+ (for me to feel pleasure now, instead of pain now, it must be that I (my proper parts) realize a different low-level institutional characterization now), it must be that there is also a functionally characterized physical difference. That is, even if the physical tokens are the same, they must be structured differently. But, now, why must this be the case? Let us run through the metaphysical necessity claim again.

Metaphysical necessity is generated, in a way that will support materialism, conditionally—that is, if $A = B$, then necessarily: $A$ is identical with $B$. Thus to get the necessity claim going, it must be true that $F = P$. Let us first consider the claim that it is the functionally characterized low-level syndrome, +F+, that is identical with +pain+. That is, let's flush out the token-identity claim that particular property instances of +pain+ are identical to low-level functional syndromes realized by tokens reductively identified with microproperties of the institutional tokens.

I argue that the essence of my particular pain now is not the functional role +F+, but the ostensively identified what it is like feature of the pain state, viz., +pain+. But then, +F+ $\neq$ +pain+, for the functional role is not played by the what it is like feature of pain (which cannot be described
in purely functional terms) by which an experiencer of pain, one aware of +pain+ identifies pain sensations as pains—although it may be that +F+-as-realized gives rise to +pain+. By "+F+-as realized" I refer to a particular realization of the institutional type by microphysical items. For +P+, "+F+-as-realized" refers to microphysical property instances of N. I use the more cumbersome locution, "+F+-as-realized", for it would seem that for materialism to be true and functionalism to be an adequate account of particular mental occurrences, talk of both functional organization and realizers is required. That is, it certainly seems that a particular (unstructured) subset of noiceptive neurons in a vat will not be a +pain+ even if those tokens are deemed necessary to an organism's feeling pain. Moreover, mere functional characterization, construed as an abstract notion, cannot itself be a +pain+ (it must be realized). The locution "+F+-as-realized" reminds us of this.

I find some uneasiness surrounding this notion of "realization". The particular +pain+ of which I am aware now (in my neck) cannot be said to be identical with +F+ nor with the items that realize +F+, but with +F+-as-realized. Thus, that it is necessary for the occurrence of +pain+ that neural (physical) items be organized in a certain way, does not amount to the identification of the micro-properties of that low-level organization with +pain+. Yet
the claim that such purely physical microproperties can realize the causal role attributed by +P+ to +P+ seems very close to identification, for why think this is possible unless the correlation between reports of felt pain and inner physical happenings is interpreted as that of identity? Again, the response: the correlation of concern is that between such reports and functional syndroms as realized. But this leads to the claim that +P+ = +P+-as-realized, and when "=" is understood as "reductively identified with", we have seen reason to deny it.

As we have seen, there is little reason to think (though lots of motivation) "+P+" refers to such functionally characterized physical property instances, for when in pain we are aware of +pain+ and despite the fact that our descriptions (first-person reports) of that of which we are aware are topic-neutral-cum-public, the object of our awareness is not topic-neutral nor is our knowledge of such objects. This marks a crucial distinction between +P+ and +P+-as-realized. Unless this distinction can be thwarted, phenomenal property instances cannot be reductively identified with functionally characterized physical property instances, and materialism suffers.

There is, then, some plausibility to the claim that one cannot be in a mental state without being in the functional state which mediates between stimulus and behavior. But
being in a functional state, on my understanding, is having a functional characterization \( F \) true of oneself. That is, I am in functional state \( F \) when I (my proper parts) realize functional role \( F \). Barring reductive identification, a natural reading of this, however, is that whenever "+\( F \)+-as-realized" is true (whenever I am in the functional state characterized by +\( F \)+), +pain+ supervenes. But now, leaning on materialist proclivities, when a particular +pain+, +P+, in an organism at a time \( t \), supervenes on the microproperties of neuroanatomical (physical) structure \( N \), which realize functional role +\( F \)+, what sense can be given to "supervenes" in +\( P \)+ supervenes on +\( F \)+ and hence, supervenes on \( N \)?

One may wish to maintain that +\( P \)+ supervenes on +\( F \)+ (and hence, \( N \)) in the sense that psychological descriptions supervene on functional-cum-physical descriptions. Thereby, one may then claim, the former (psychological language) is a way of interpreting at a molar level the significance of lower-level organization. But this interpretation is ontologically questionable since it would (at best) only apply to third-person accounts, where correlations are assumed to have been established. However, even given such correlations (between lower-level functional-cum-microphysical happenings and first-person reports) it does not follow that the ontologically charged claim that +\( P \)+ = +\( F \)+-as-realized is true. For the latter is an interpretation
of such correlations; an interpretation, I maintain, tainted by ontological considerations raised above (see chs. 1 and 2).

Without reduction, to say \( +P^+ \) supervenes on \( +F^+ \) is to say \( +F^+ \) (as realized) causes \( +P^+ \) or \( +P^+ \) emerges. Again, since supervenience is transitive, \( +P^+ \) must supervene on \( N \), which is to say \( N \) causes \( +P^+ \) (in virtue of realizing \( +F^+ \) perhaps) or, again, \( +P^+ \) emerges. (Yet, if the supervenience relation is causal, does \( N \) cause \( +F^+ \)? Perhaps, I leave that question aside.) Saying, as does Kim, that the relationship between \( +P^+ \) and \( N \) (foregoing reductive identity) "is not happily considered as a causal relation, although it is an important variation of what may broadly be termed a "determinative relation"" succeeds only in naming the problem; for what a "determinative relation" is, is no clearer than what a supervenient relation is—it is only clear that neither is happily considered causal.

It may still be thought that I have yet to show that \( +F^+\text{-as-realized} \neq +P^+ \). This shows how we must construe supervenience. To what does "+F+ as-realized" refer if not the realization of \( +F^+ \) by microproperties of \( N \) upon which \( +P^+ \) supervenes? But this (taking supervenience as non-reductive) leads to a trivial identity claim in that the realization of \( +F^+ \) by \( N \) is what accounts for the supervenience (emergent arising) of \( +P^+ \). Thus, \( +F^+\text{-as-realized} \)
(my being in functional state \( +F+ \) at \( t \)) = \( +P+ \), trivially, because (so to speak) \( +F+ \) supervenes on the left-hand side of "=" and, of course, \( +P+ = +P+ \). What is trivial in this identity claim is that "=" is not construed as reductive identification, but biconditional correlation— it only says that whenever \( +F+ \) is realized, \( +P+ \) emerges—in the nasty sense of emergence. To make sense of the identity claim, we must construe supervenience as reduction of (phenomenal) property instances of \( +\text{pain}+ \) to functional-cum-physical property instances. I argue that this reduction fails.

Once again, consider Jones, who knows the complete functional/physical characterization of \( +\text{pain}+ \), but is himself congenitally unable to feel \( +\text{pain}+ \); there is something Jones fails to know about (the concept of) pain, namely, what it is like to be in the pain state. Jones' knowing that \( +F+ \)-as-realized = \( +P+ \) only amounts to his knowing that the sentence "\( +F+ \)-as-realized = \( +P+ \)" is true, it does not enable Jones to know about \( +\text{pain}+ \), as he would if he were aware of \( +\text{pain}+ \).

For, if functionalism is true, Jones has no \( N \) (physical guise) which realizes \( +F+ \). Should Jones come to have the appropriate \( N \) realizing \( +F+ \), then he would, ceteris paribus, know about \( +\text{pain}+ \). Why? Because Jones, by having the appropriate \( N \) which realizes \( +F+ \) (by being in functional state \( +F+ \)) would have an awareness of the \( +\text{pain}+ \) which supervenes. Since it would seem that one is in the pain
state if, and only if one is aware of +pain+, we must (again) examine the incorrigibility thesis, for Lycan\textsuperscript{45} maintains that this claim is equivalent to the incorrigibility thesis which, given Armstrong's arguments,\textsuperscript{46} is suspect.

Before turning to the incorrigibility issue, however, let me summarize. The above explication of the functionalist strategy has, no doubt, appeared redundant. In a sense it is, because the account given for particular occurrences of +pain+ amounts to reiterating the homunctionalist strategy for being in pain. To be in the pain state is to have a certain functional characterization true of oneself. Suppose we agree that one is in the pain state if, and only if one realizes a functional syndrome. This leaves it open that realizing a functional syndrome merely gives rise to +pain+ as an emergent phenomenon. In order to defuse this emergence, +pain+ itself must be given a like treatment. It is this latter, I have argued, that fails. Functionalism may indeed provide us with necessary and sufficient (nomological) conditions for being in the pain state, but it does not adequately account for what happens once we are in that state---the emergence of +pain+. At best it can tell us that we will feel (be aware of) the +pain+ that emerges when certain microproperties are arranged in certain ways. But what is required of materialism is that these property instances be reductively identified with purely functional-
cum-physical microproperties of the tokens realizing the functionally characterized states. It fails in this because what is essential to being aware of +pain+, the +pain+ itself, the what it is like feature of pain states (the quality-object), being ostensive in character, does not succumb to functional/physical treatment.

Now, incorrigibility again. Lycan cites Armstrong, who argues that a brain technician of the future might have correlations establishing that +P+ is experienced when and only when one is in a functional state realized by one or another of a subset of noiceptive neurons, N₀, . . . , Nₙ. Imagine, however, that Smith (truthfully) reports he experiences +P+, but is in none of the functionally characterized N₀, . . . , Nₙ. On the basis of his evidence, the technician concludes (and claims it ought to be concluded) that Smith has made a mistake. (Let us assume that Smith's belief does not have a statement as its object.) Assuming that Smith (an accomplished sensation-language user) is telling the truth, attending to his mental goings-on, and has not made a verbal mistake (has not mistaken +fear+ for +pain+), it would follow that his belief about his psychological state, that it has the property +P+, does not entail Smith's being in a psychological state with property +P+. That is, Smith may believe he experiences +P+ when he does not; his report is not incorrigible. Armstrong assumes, however,
that we could have evidence which shows that Smith believes he is experiencing +P+ when he is not, that is, that it is conceivable that a conflict between Smith's sincere report and a machine reading will arise. This tack will work only if it is not more plausible to construe this evidence in such a way that it renders more plausible the conclusion that Smith does not believe he is experiencing +P+. Armstrong's argument fails here.

Although it is conceivable (given current neurophysiological knowledge) that such functional-cum-physical-mental correlations will be discovered, it is not obvious (given an understanding of how such correlations will be discovered) that such a conflict will arise. Again, we must bear in mind that the evidence Armstrong relies on must render plausible Smith's not being in pain, while not rendering it plausible that Smith does not believe he is experiencing +P+. Consider Pappas:

If it is now reasonable to think that laws relating sensation and neural events will be built up in the future, then it is also reasonable to think, now, that EEG-conflicts will not arise. This is simply because in order for these law-like correlations to get established in the first place, there must be extremely reliable correlations holding between subjects' reports and machine readings; and an EEG-conflict would simply be a breach of these correlations. And surely it is reasonable to hold, now, that future neurophysiological science will consist, in part, of laws relating experiences of specific sensations and specific neural events.48
Note that the claim here is not that a counterexample will denigrate the correlations already (let us assume) established, but that no (future) conflicts will arise. If we have good reason for holding that none will arise, Armstrong's reason for claiming Smith's report is not incorrigible will vanish. We must remember that Armstrong's case is merely hypothetical (a mere logical possibility?), thus should it be maintained that "conceivable" in "it is conceivable that Smith's belief that he experiences +P+ does not [in this hypothetical case] entail that he experiences +P+ (that he is in pain)" is to be understood as logically possible, then there is a retort. It is (also) logically possible that when Smith does not experience +P+ (when he is not in pain), it logically follows that Smith does not believe that he experiences +P+--logical possibility slices both ways. That is, given that (in our hypothetical case) the (future) technician is completely justified in maintaining the correlation in Smith between +P+ and the functionally characterized N, it is as plausible to hold (now) that if (in the hypothetical case) Smith is not experiencing +P+, he will not believe he is as it is plausible to hold that he believes he is experiencing +P+, but is not in pain. If it is logically possible that a conflict will arise, it is also logically possible that none will arise. But this is not a draw; since there is an argument to the effect that
such a conflict will not arise, Armstrong's claim is suspect—especially when we add the intuitively plausible belief that one is in pain if, and only if one believes he is experiencing $+P_+$. Armstrong's case does nothing to denigrate this latter conviction. In short, why think the hypothetical case will come about in such a way to establish the refutation of incorrigibility?

(Moreover, as we have seen, Armstrong's claim is even less plausible when we replace "believes he experiences $+P_+$" with "is aware (de re) of $+P_+$". For consider: Smith may be aware of $+P_+$ when he is not in pain. If so, "Smith is in pain if, and only if Smith is aware of $+P_+$" is not necessarily true, that is, it is logically possible that Smith is aware of $+P_+$ when not in pain. Perhaps Hume's "law", that no two distinct existences are logically connected, supports this. Is it logically necessary: if $x$ is red, then $x$ is colored? Perhaps, I don't know; but I would maintain (at least) that it is conceptually analytic. Where a statement is conceptually analytic if it expresses a connection constitutive of the connected concepts, I would (at least) maintain it is conceptually analytic: Smith is aware of $+P_+$ if, and only if Smith is in pain. Barring a "conceptual revolution", this gives incorrigibility enough substance to allow one to deny that it is conceivable that Smith's being aware of $+P_+$ does not entail that he is in pain. I take
it that conflicts that arise now "favor" Smith, not the technician.)

If the above reasoning is correct thus far, it follows that the type of law (or necessity) involved in \( L \) cannot successfully be understood as metaphysical. Nor, it seems, will it do to understand it in a causal way, for then we would have:

For each psychological state \( M \) there is a functional state \( F \) such that, as a matter of causal law, a state of type \( M \) occurs in an organism at \( t \) just in case a state of type \( F \) is realized by the organism at \( t \).

But then, though we have a causal explanation of the appearance of \(+pain+, \) we are (leaning on materialist proclivities) back to the problems of correlation and psychophysical causation or emergence. Barring an identification of a particular occasion of \(+pain+\) with a functionally characterized neural structure, \(+pain+\) appears as an epiphenomenon with no causal powers of its own. That only physical structures will realize the functional properties that define mental states seems little different from Descartes. And if supervenience is not causal, it would seem we are headed for a \(+P+\text{-}N\) parallelism. Now I turn to Kim and problems of psychophysical causation.
Recurring Problems

Regarding the notion of supervenience, Kim writes:

A family $M$ of properties is supervenient upon a family $N$ of properties with respect to domain $D$ just in case necessarily, for each property $P$ in $M$ and each object $x$ in $D$ such that $x$ has $P$, there exists a property $Q$ in $N$ such that $x$ has $Q$ and any object $y$ in $D$ which has $Q$ also has $P$.\(^{50}\)

Where $M = \text{the family of mental (phenomenal) properties}$ and $N = \text{the family of functional properties}$, it would seem, given materialist concerns, that $D = \text{neural (physical) objects}$. Note that token-identity appears built in, for the properties of $M$ and $N$ are properties of the same objects. Thus, my particular $\text{+pain+}$, $+P_4+$, has a sufficient condition in the particular neurological correlate, $N_{40}$, that realizes, say, $F_{14}$ (a functional property of $N$). Kim maintains, since he recognizes the force of the pro-qualia intuition, that the supervenience relation is not one of identity, but of nomological equivalence—we are dealing with nomological necessity. He notes:

\[\ldots\text{ if every property } P \text{ in the family of properties } M \text{ has a nomological equivalent in the family } N, \text{ that is, there is a bi-conditional correlation (with a suitable modal force) relating each property in } M \text{ with some property in } N, \text{ then } M \text{ is supervenient upon } N.\]^{51}
(For Kim, the "suitable modal force" varies with the particular occasion of use; I will assume here that it is nomological necessity.) Where we are talking of instantiated properties, this view leads to the problems encountered earlier (see Chapter 1), for it is not clear what the ontological status of any particular phenomenal (emergent/supervenient) property instance amounts to. Thus, although we have token-identity of individuals, we may have a dualism of property instances (not individuals?). Either these particular (instantiated) properties are causally efficacious or they are not. Kim's view is that they are not (but see below), although his wording indicates some confusion. He states that a particular +pain+ is a "supervenient cause" of a groan (and, I assume, avoidance behavior); this is understood to mean that the +pain+'s "causal role is dependent on that of its neural correlate which has a direct causal role with respect to the groan." (In our amended functionalist version, the neural correlate is N_{40}, which realizes F_{14}). To grasp the nature of the dependence of the supervenient causal role on the direct causal role, consider a revised version of our earlier diagram.

\[ \text{+P}_4 \rightarrow \text{G} \]

I ?

N_{40} \quad N_{400}
Letting the broken arrow represent the notion of "supervenient cause" and the solid that of "direct cause", we see that our earlier problem arises. There is a direct causal path (link) through the neurological items to G (the groan). Hence, this direct path represents the robust causal role and is played by physical items. The only reason for holding that the broken arrow represents a supervenient causal role appears due to the biconditional (or sufficiency) relation holding between +P+ and N\(^q\).

Yet this is not robust causation since +P\(+4\) is a phenomenal (non-physical) property instance. (If the supervenient cause is thought to be robust, we have interactionist worries, for we have no account of how non-physical items could exert forces on physical items so as to cause G.)

"Thus, the pain is not fully comparable to its neural correlate with regard to the causal role vis-à-vis the groan; in an important sense the pain has no independent causal power. . . ."\(^{54}\) (For my understanding of this "important sense", see chapter 1 above.) This matches our earlier conclusion. With regard to giving an account of the observed correlation, Kim offers little cash:

Both the pain and the groan are determined by the brain activity that forms the supervenience base of the pain: the relation of this brain activity to the pain is that of supervenient determination, and the relation of the brain activity to the groan is that of causation.\(^{55}\)
Two questions arise: (1) What sense can be made of "supervenient determination" that will help the materialist? and (2) Isn't a particular +pain+, e.g., +P+ a mere epiphenomenon? With regard to (1), Kim resists calling the supervenient relation that of identity because "[i]n the mind-body case there are some persuasive arguments to show that mental (especially phenomenal) properties must be distinguished from physical or neurological [or, I would add, topic-neutral] properties." And, if my arguments above and in following chapters have plausibility, he is wise to avoid the identity account. Although, as I have indicated above, Lycan would appear to favor the identity interpretation of the supervenient relation, it is not clear how he can and still give force to the pro-qualia intuition that there are particular +pains+ of which we are aware. Thus, (1) remains unanswered, except to say that there are certain law-like regularities, namely, +pain+ wherever certain neurological properties (or the realization of certain functional roles).

Concerning (2), Kim does little more than recognize the difficulty. We are told that his "'Supervenient dualism'... bears a family resemblance with epiphenomenalism, but I believe that this is not to the discredit of the theory." Why he believes this I do not understand, for he simply reiterates his belief that macro-phenomena are, in general,
supervenient on micro-processes where the "causal roles" of the former are explicated by the latter. Question (2) is not happily resolved.

Not adopting the identity solution and ruling out epiphenomenalism, though, leads us to consider the assumption that qualia are causally efficacious. But this raises its own (well known) puzzles. It is interesting to simply note that the scientific world (or at least part of it) is embroiled with the same conundrum. Developing an argument based on our commonsense views that +pains+ are causes, as well as the evolutionary view that +pains+ have survival value and thus, causal usefulness, R. W. Sperry sets forth his view of mental supervenience:

It is the idea, in brief, that conscious phenomena as emergent functional properties of brain processing exert an active causal role as causal determinants in shaping the flow patterns of cerebral excitation.58

The subjective qualitites are recognized to be real and causal in their own right, as subjectively experienced, and to be of very different quality from the neural, molecular and other material components of which they are built.59

Let us return to question (1). J. L. Mackie argues for a view of causation such that

... if on a particular occasion A's doing X is causally related to B's doing Y, and if they had not been so related but things
had otherwise been as far as possible as they were, A would still have been doing X but B would (or might) not have been doing Y, then A's doing X is conditionally and causally prior to B's doing Y.60

Accepting this, as I do here, +P+4, as an epiphenomenon, would be construed thus: if the supervenient link between +P+4 and N40 had been missing, but things had otherwise been as far as possible the same, N40 would have been as it is but +P+4 would not have been there. Thus, +P+4 would be causally idle, for if the supervenient link had been missing, etc., the lack of +P+4 would not have prevented the occurrence of (say) the groan. Mackie's move is to claim that according to the physical-chemical laws that constitute the framework within which evolution takes place, the supervenient link between +P+4 and N40 could not have been lacking--"neither item could have occurred without the other". Mackie reasons:

If mental features had been causally idle, how could they have been naturally selected? The reply is that if these features were linked to their neurophysiological basis by laws of the framework within which evolution occurred, and this basis gave rise to actions that were useful for survival, these features would automatically be selected along with their basis: evolution simply should not have developed the basis without the consciousness.61

Although interesting and certainly a possibility, one's first response is that +P+4 is not causally operative itself,
it only "inherits" a causal role in the evolutionary story. But then these features (e.g., +P^4) do not themselves appear to have been naturally selected. Let me work this out a bit. First, everything hinges on the claim that +P^4 and N_{40} are causally inseparable, not merely in the sense that they are as a matter of fact inseparable, but that it could not have been otherwise. But surely N_{40} could (logically) have evolved without +P^4, even if in fact it did not—Mackie offers no reason for thinking otherwise. And failure to resolve this would seem to leave the evolution of consciousness a mystery, in that there is no reason given why there should be a nomic connection between +P^4 and N_{40}. It seems, moreover, more correct to say that +P^4 was naturally selected by accident (if this makes any sense), for on Mackie's view it would appear that +P^4 was not naturally selected because of its causal role, but rather that +P^4 has a causal role because it was selected. But then, as mentioned in Chapter 2, +pain+ appears, in a sense, gratuitous—a cruel accident perpetrated by Mother Nature. Finally, even if we waive these difficulties and maintain that the question of why there are +pains+ makes no more sense than the question of the purpose of life, we are stuck with epiphenomena, for +pains+, unlike chins, are not mere pleiotropisms, but are elements of being. In fact, Mackie's view seems (as has been argued above) a plausible argument for epiphenomenalism as opposed to
interactionism—still, however, materialism has not benefitted, at least not directly.

Inverted Qualia--Resolution (Of A Sort)

Finally, let's return to Lycan's way with inverted qualia. It should be clear that Lycan's move is to claim that we do type-identify our own more qualitative psychological states (on the basis of felt simplicity) and corresponding to this ability will be a discoverable (and programmable functional difference between, e.g., (occasions of) +red+ and +green+. Thus, he suggests that a functional difference may be found between Sid and Amy if we but, as Sober says, look inside. We should note that the possibility (other than mere logical possibility) of inverted spectrum (qualia) is conceded only relative to skin-out behavior.

In setting the problem above, I claimed that Sid and Amy are behaviorally indistinguishable and that "even given future neurological research, that Sid and Amy are functionally indistinguishable." Thus Lycan is not conceding that inverted spectrum is possible (other than merely logically possible) relative to functional profile. What this suggests is that future neurological research may dissolve the asymmetry problem; this move also brings into play the distinction between functionalism and behaviorism. In effect what is being claimed is that it is our ignorance of
lower levels of functional organization, where structural
differences may account for the qualitative differences
between Sid and Amy, that makes inverted qualia objections
seem so intractable for the functionalist. Lycan seems
correct in questioning the assumption of those who push the
inverted spectrum problem at the functionalist's doorstep.

*If* the token-identity claim were true, Sidney Shoemaker's
similar response would settle the matter:

> In principle we can go hunting in the
> physiology of a creature for a pair of
> physical properties which play the
> functional role of qualia, which are
> involved only in the perception of blue
> and yellow, and which are such that, as
> long as one of them characterizes percep-
> tions of blue the other can characterize
> only perceptions of yellow, and vice
> versa. . .65

Switching only the colors, we see how this applies to
the present case, for once we find the physical properties
which are involved in the perception of *red* and *green*,
we need only look inside Sid and Amy (given future neuro-
logical science) and see if these properties match, that is,
see if there are physical properties in Sid and Amy that
play the same functional role. How are such properties to
be located? Here Lycan's suggestion seems apt; we must
locate the level of organization at which a difference makes
a (relevant) difference. Since a great many of the physical
properties of Sid and Amy will differ, focusing on the
relevant level is crucial. If we should discover (via correlations) that the relevant level is the neuronal level, we look for the difference there. Given the plausibility of this multi-leveled analysis, we must simply wait and see what future internal investigations turn up in the way of complex structures which may serve to distinguish +red+ in Sid from +green+ in Amy (here the top-down meets the bottom-up strategy). Since inverted qualia is only a logical possibility, the logical possibility of there being the relevant correlations is all Lycan needs.

Lycan, then, holds that (materialistic) functionalism will be able to handle inverted qualia as long as empirical research turns up sufficient structural complexity enabling correlations of the right sort. Moreover, given Dennett's pain data, he believes there is good reason to think it will. There is considerable plausibility to this way with inverted qualia. For one thing, it does seem plausible that corresponding to my feeling pain and your feeling pain is a functional similarity. It may also be that between Sid and Amy future brain scientists will discover in structure a functional difference. This move also avoids the (false) dilemma Churchland and Churchland attempted to construct, for what all pains have in common need not be either an objective intrinsic physical nature or a non-objective intrinsic non-physical nature. Instead they may share a
functional essence: having a φ-er in the appropriate state. It also seems more plausible than Dennett’s outright banishment of qualia, for it recognizes (to some extent) the strength of the pro-qualia intuition by giving +pain+ a level of reality (albeit a very low-level).

But, of course, what is given a level of reality is +pain+ construed as a universal (functional type)—what of the particular property instance of the pain state, viz., +pain+ of which I am aware now? (Or, what seems even worse, "+pain+" is given a level of descriptive reality, in that psychological language supervenes on functional-cum-physical-language.) Even given the relevant correlations, the ontological issue remains unresolved: how are we to make sense of the claim that a particular property instance, +pain+, is (reductively identical to) the realization by physical microproperties of a functional (homunfunctional) state? I agree that if an entity (e.g., a sponge) lacks some (unspecifed) degree of complexity, then it (probably) cannot feel +pain+. But when an entity (system) does have the required complexity and does feel +pain+, what is the particular +pain+? To be hold that functional goings-on are both necessary and sufficient for the occurrence of a particular occurrence of +pain+ appears to leave us begging for a plausible analysis regarding that of which we are all so well aware—+pain+. My worry is that even if we discover how
qualia come about and what they are correlated with, unless we can say what they are (and all of what they are) in functional/physical terms, unless we can provide acceptable analyses which comport with the phenomenological data, we are left with Dennett's worry—nasty emergent phenomenon. We see, then, that Lycan's way with these difficulties leads to the same arcanum arcanorum. Materialism is yet to be home free, for my particular +pain+ cannot plausibly be reductively identified with a low-level functional organization of purely microproperties of physical states, for felt pains cannot be analysed as functional properties (cf. the discussion of topic neutral properties in chapter 1).

Note, however, that given the above explication of functionalism, certain objections based on the notion of absent qualia, lose much of their intuitive bite. Consider Searle's understanding and criticism of functionalism:

For a human to be in pain or some other mental state is just to instantiate a certain sort of functional organization ... with the right inputs and outputs.

Any system such as a robot that was functionally equivalent to a human, i.e., that had the right organization and the right input and output relations, would also have to be in pain in virtue of those facts.

Searle's objection seems to hinge on the following: if functionalism were true,
We could, e.g., make the system out of windmills, old beer cans, and paper clips, and according to organizational functionalism, any such system must be in pain. Against this I am arguing that from everything we know about human physiology, it seems incredible to suppose that these other systems could be in pain, much less that they must be in pain.68

This is close to a "howler", for the enlightened functionalist would not doubt agree that such systems could not be in pain, if "could" is interpreted in the sense of physically possible. But the functionalist is not requiring the physically impossible, but only maintaining that if such systems did realize the appropriate functional syndromes, then they would be capable of feeling pain. For functionalism, as understood herein, would claim that such systems, if they did have the required complexity, if they did realize the pain-functional syndrome, including the low-level aspect, they would become, it seems, "perfect" sentient simulators. That beer cans and paper clips appropriately strung together would not be such sentient systems only seems intuitively plausible because it is not plausible to think they could (physically) have the requisite (micro-) properties for low-level realization (for one thing they do not seem to be teleological systems). But if they did they would be systems capable of feeling pain, after all, we realize (via our proper parts) such syndromes and we feel
pain. What do we know about human psychology according to which "it seems incredible to suppose that these other systems could be in pain"? For one thing, human physiology seems much more complex and (teleologically) organized.69

My worry is that if such "perfect" sentient simulators did feel pain (were aware of +pain+), we would then have a "mind-sentient simulator" problem. Once again, +pain+ would emerge—if we are to continue speaking of +pain+ as phenomenologically understood. That is, I am not of the bizarre opinion that human beings are not composed of physical stuff, but only that the physical stuff gives rise to phenomenological property instances which are emergent. (OK, so I'm not of the opinion that human beings are composed of purely physical stuff.) Moreover, due to the ostensive nature of +pain+, I do not see how it can be given a functional characterization which is constitutive of (reductively) identical to +pain+, thereby avoiding such nasty emergence. Mysterious, yes; bizarre, no—that's where the argument leads unless the data is denied or given an implausible interpretation.
When light from a star reaches our eyes, it is the end of a long train of events which can be traced by physics. But the sensation of sight is the only event in the whole series about which we can say anything not purely abstract and mathematical. A blind man might know all physics, but never the sensation of seeing. A knowledge that things are pleasant or unpleasant is not physics. Hence it is clear that there is knowledge which is not included in physical science—a knowledge of our own mental sensations.\(^1\)

---Sir William Dampier, 1929

Functionalism, Token-Identity, The Role of Experience——An Introduction

Earlier I suggested that phenomenal properties are distinct from physical properties in the sense that "pain" is an ostensive term whose descriptive/reportive role cannot be filled by the non-ostensive terms having to do with physiological (physical) or functional (topic-neutral) facts or properties. I want now to elaborate upon this claim as I think it leads to the phenomenal property objection in its strongest form and casts serious doubt on the tenability of materialism. I begin by noting how language with empirical content is dependent upon our experience of the world. The link between language and the empirical world
comes via ostensive "definition", where reference to certain items as presented in experience is not mediated by theoretical concerns. Although one may find this a contentious claim, I find it incontestable if we are to avoid the view that our theories can change the world and not just some views of some parts of it.

Our link to the empirical world, our ability to employ and interpret language with empirical content, our ability to conceive of the empirical world necessarily presupposes, depends on, our experiencing that world. Any theory (language with which we communicate) about the empirical world must take into account the objects of experience in terms of which our experiences are identified and discriminated, in terms of which our empirical concepts are learned and to which they refer. When (under normal circumstances) I touch a red-hot iron, I identify my psychological state at $t$ as a pain state (I know "pain" can be correctly predicated of me at $t$) in virtue of the painfulness of my experience and the fact that I am an apt language user. It will not do to employ a form of "conceptual relativism" and claim that all the content of my experience is subject to revision on theoretical grounds, for there is a world independent of our theories and it can be known in a theory-independent way. The painfulness of my experience cannot be made to vanish by employment of a theory种种, what I experience when I
touch a red-hot iron, by any other name is still +pain+.

Our access to the empirical world is via our experience; since this world is pre-theoretical, any two theories about this world must have something in common if they are to be theories of the same thing. The pre-theoretical world limits our theories, our theories do not limit the empirical world, else we could change this world with our theories and, for instance, theorize felt pains away—we cannot pace Dennett. Facts are not little theories, not big theories facts—at least not with regard to the empirical world upon which experience impinges. Thus it seems that any two theories of pain (that which we experience) must have an element in common: our pre-theoretical experience of the hurtfulness which is identified ostensively (albeit not in utter privacy, see below). Since we recognize this element, again and again, we name it, and what "pain" in theory T₁ denotes must be the same as what "pain" in theory T₂ denotes—the what it is like feature of our experience; +pains+ are not like demons or witches.

But, given that any two theories of pain must have an element in common, why think it must be +pain+? Perhaps, as Lycan suggests, "pain" is a natural-kind term and what theories of it have in common is reference to its functional subroutine. Still, +pain+ cannot simply be eliminated à la Dennett; there must be a "bridge" from "+pain+" in our
"folk" theory to "pain" in our scientific theory. That is, there must be a plausible reduction, for to say "what "pain+" really refers to is the functional subroutine (or its realization), not a private phenomenal object," needs backing if we are to avoid theft over toil. Lacking such a reduction, there is little reason to think the scientific theory a theory of our experience of the feel of pain; we would be justified in claiming something left out. I argue that there is reason to believe that "pain+" does not refer to functional subroutine. Hence, any theory of pain is constrained in that the role of experience in identifying pain must be taken over by the items which the theory claims are identical with (or to which are reduced) the objects of experience. That there are, prima facie, such "objects" of experience becomes palatable when we realize that we recognize or discriminate our experiences with regard to the qualities ("objects") of our experience.

Thus, the question is whether physical or functional items can play the role experience plays in accounting for our knowledge of what it is like to experience +pain+. That is, can physical/functional items serve as the items by which we identify and discriminate a sensation as, say, a pain sensation rather than a tickle sensation in our own case? I take this question as pointing out an adequacy condition for any theory of pain. Moreover, these items must also play the
role felt pain plays in forming our attitudes toward fellow pain feelers. My sympathetic understanding (perhaps something akin to Dilthey's verstehen) of Morgan's suffering is due to my ability to imagine myself as having touched the red-hot iron—I realize that I feel pain, not because I cry out, but that I cry out because I feel pain. Nor does the claim that there can be no purely ostensive language jettison the idea of ostensive identification. The ostensive learning of "pain" does require public stage-setting (see below), but it does not follow that what "pain" refers to is theory-dependent (except in the way that any word is dependent on a language for its home).

Before turning to the private language argument, let us rehearse the structure of the functionalist picture. A brief and (very) schematic look at how a functional characterization of pain may be arrived at will provide further elucidation of the functionalist claim. Since present concern is with psychofunctionalism, I shall attempt a proto-scientific account remembering, however, that we are trying to bridge the gap between the Manifest and Scientific Image. Commonsense and the data of introspection would have us claim that the cause of pain-behavior (or its accompaniment) is the feeling of pain—I withdraw my hand from the red-hot iron because it hurts. Employing what Dennett would label an a prioristic top-down strategy, let us investigate
the logical/causal constraints involved in formulating a functional answer to the question: "What would have to occur in Jones for him to engage in pain-behavior?" (Again, I am interested only in the barest outline, the skeleton of this process.)

To begin let us hypothesize an external cause (stimulus) for the overt pain-behavior (e.g., withdrawal, grimacing, reports of pain, etc.) typically thought to result from felt pain. (I shall ignore purely internal causes and only be concerned with "physical" pain. Also, let us ignore the question of just what constitutes pain-behavior and assume it is what we typically think it is.) Having learned from the behaviorists' failures that simple input-output explanation in terms of overt behavior is neither necessary nor sufficient for the occurrence of felt pain, we continue hypothesizing causes—in effect developing a causal chain or functional profile which takes us "inside" the organism. Assuming (perhaps finding) correlations between the overt behavior and internal mechanisms, we posit internal causes of the overt behavior such that where the overt behavior is (is not) manifested it can still be maintained that Jones is not (is) in +pain+. This can be accomplished as long as the functionalist can give a functional/causal account of why the overt behavior did not occur, but felt pain did. A similar account must be presented for why overt pain-behavior
did occur, but felt pain did not. We see that this talk of functional/causal role brings us to the notion of "realization," as Shoemaker remarks:

A physical state or property realizes a functional state or property in a particular creature if in the workings of that creature it plays the "causal role" definitive of that functional state or property, i.e., if it interacts causally in the required ways with inputs, outputs, and other internal states of the creature. Given the sense in which they are functionally definable, qualia too can be said to have physical realizations, which in principle we could discover by physiological investigations.5

Still, we may wonder about these physical realizations; are they what particular +pains+ are? Simplifying even more, it may be claimed that $F_4$ is the institutional type identified with withdrawal behavior (of the hand from the hot iron), $F_8$ the institutional type identified with exclaiming "Damn, that hurts!", $F_{20}$ identified with changed blood pressure, and so on. Eventually the question arises of what to identify with the feel of pain. Assuming that the functionalist wishes to appease commonsense and common experience, it would seem natural to identify felt pain with the functionally characterized inner activity, $E_{14}$, which (as realized) is eventually causally responsible for typical pain-behavior. It can then be claimed that Jones feels pain even if he is a super Spartan and does not engage in,
say, withdrawal behavior—there will be a functional/causal explanation of this (atypical) behavior despite the realization of $F_{14}$. Likewise for Jones the actor; he will not be attributed felt pain (by the knowing functionalist), despite his performance, because the functional/causal explanation rules out the realization of $F_{14}$.

Notice that things begin to take on a dispositional air; $F_{14}$ is a low-level institutional type which, when realized (in the appropriate context), is causally responsible for pain-behavior (overt and/or covert). (Also note that I am assuming that the functional significance of pain attribution is some way tied to behavior.6 But now, if $F_{14}$ is realized by physical token $N_{40}$, then $N_{40}$ is what, in Jones at $t$, caused his pain-behavior. What principled reason is there for claiming that a particular felt pain at $t$, $+P_{4}$, is identical with $F_{14}$ as realized by $N_{40}$? Although it may be denied that the requisite internal activity and outward behavior associated with felt pain in Jones at $t$ would have occurred without $N_{40}$ realizing $F_{14}$, that should not matter unless $+P_{4}$ is identified with $F_{14}$ (as realized) and/or outward behavior itself. Yet the inner activity and outward behavior only seem to clue us in our attribution of felt pain to Jones at $t$, and $N_{40}$ serves only to ground that activity/behavior. We are back to an identity claim in which $+P_{4} = F_{14}$ (as realized by $N_{40}$); for only the truth of
this identity can render functionalism plausible as a materialist theory of mind. So, for the functionalist theory to be a theory of pain (+pain+) at \( t \), there must be reductive identification of +P_4 with \( F_{14} \) (as realized by \( N_{40} \)).

It seems, then, that the functionalist is telling us that all there is to +pain+ (in Jones at \( t \)) is captured by the functionalist account: for Jones to experience +pain+ at \( t \) is for Jones at \( t \) to realize a specific functional syndrome. If he realizes that syndrome, he cannot fail to experience +pain+; the necessity here is due to interpreting the correlation between realization of \( F_{14} \) by \( N_{40} \) and the occurrence of +pain+ as due to identity: \( N_{40} \) realizing \( F_{14} \) (in Jones at \( t \)) is constitutive of +pain+ (in Jones at \( t \))— that's all there is. Returning to our adequacy condition for a theory of pain, we find that it must be possible that (in our own case) what we are aware of, what we use to discriminate our state as a pain state, must be how \( N_{40} \) appears to us—in fact, not only must it be possible, but it must be that this is what we are really doing when we identify and discriminate our own states (appearances to the contrary). But this brings double aspect worries, for we must now think of our awareness of \( N_{40} \) as awareness of +pain+ (the emphasis is not that we must think of our awareness of +pain+ as awareness of \( N_{40} \)). That is, we must say it is our awareness of \( N_{40} \) under the guise of +pain+ by which we
identify and discriminate our sensations. But the guise (appearance) cannot be eliminated (cf. the discussion of topic-neutrality); why not think of the identification, $+P+4 = F_{14} \text{ as-realized-by-}N_{40}$, as "going from right to left", as imbuing the physical aspect of the functional story with mental features? One apparent way out of this unhappy result is by denying the appearances status as elements of being. The Private Language Argument (PLA), considered in this chapter, and the Adverbial Theory of sensations, considered in the next chapter, are often thought to solve this difficulty—-it will be argued they do not.

It may appear that there is something fundamentally incoherent about the alleged identity of institutional types (as realized) with $+\text{pain}+$. For if we identify $+P+4$ with $F_{14}$ and in Jones $F_{14}$ is realized by $N_{40}$, but in Smith by $N_{44}$, then we have the apparent result that $+P+4 = F_{14} \text{ as-realized-by-}N_{40}$ and $+P+4 = F_{14} \text{ as-realized-by-}N_{44}$. But how can one thing be identical with two things not identical with each other? The intuitive response that $+\text{pain}+ \text{ in} \text{ Jones}$ is identical with realization by one neural item and $+\text{pain}+ \text{ in Smith}$ with realization by another neural item is the correct response for the functionalist; for what we have are two numerically distinct $+\text{pains}+$. What makes them both pains has nothing to do with the intrinsic nature of the neural items that realize the functional syndromes, but is
instead due to the functional syndrome realized by those neural items. Since both $N_{44}$ and $N_{40}$ realize $F_{14}$, both have functional properties in common and these properties are what make both pains. Or so the functionalist would have it. I argue that particular occurrences of felt pain are not functional properties of (nor identical to) neural (physical) items. Functional (topic-neutral) properties cannot serve as the items by which we identify and discriminate our own sensations as sensations.

**Introspection as Brain Scanning**

The--at present--utopian kind of physical knowledge can give an account of introspection, self-knowledge, avowals, etc. We may (somewhat speculatively) say that some part of the cerebral cortex "scans" the processes in some other part of the brain. And since the scanning part would be the one connected to the motoric nerves of the speech organs, we can thus sketch--at least by way of a "promissory note"--what the scientific account of private mental states (and their avowals) would be like. It is this sort of speculation that makes (especially the Australian) materialists so confident that a physicalistic central state theory of mind is possible, and that it need not be incomplete. Indeed, in the frame of intersubjective science nothing need be left out--except the "feel" of the rawfeels. 8

Much of what follows depends on taking the data of introspection seriously and if token-identity between mental items and physical/functional items is not plausible, then the doctrine that consciousness is awareness of something
internal and physical—one part of the brain scanning another—is hopeless. If a particular felt pain is not identical with an institutional type (as realized), then it is not clear how brain scanning can be an analogue of conscious awareness. For lacking this identification, a particular felt pain is but a functional syndrome (type) which, when realized, manifests certain behavior. On this construal, however, felt pains (as functional items) are construed dispositionally; for Jones to experience +pain+ is for Jones to be in a dispositional state, which is to say that certain conditionals and counterfactuals are true of Jones in virtue of physical items realizing functional characterizations. This does not appear a happy conclusion, for unless a disposition is identified with certain causal properties (not causal powers) of underlying structure, felt pains are but abstract, constructed entities.

Where the disposition is not identified with its causal basis, the mental items no longer appear to be concrete internal happenings. Thus introspection as awareness of an internal (physical) state is no longer a plausible analogue of, say, my awareness of +pain+; for awareness of +pain+ (construed as a disposition) is awareness of an abstract item. It is awareness of a physical item which, if certain conditionals and counterfactuals are true, is a +pain+. But how could I identify my internal physical state as a
pain state (which I do) without also knowing further internal and external environmental relationships which satisfy these conditionals and counterfactuals? My awareness of +pain+ provides me with (non-topic-neutral) knowledge of my internal state (that it is a pain state) without my knowing about further abstract conditions satisfying these conditionals and counterfactuals. As Armstrong admits:

... the parallel between introspective awareness and the tactual perception of physical pressure would fail if the latter were no more than an inference ... from independently given sensations of pressure.10

Thus the notion that we are aware of mental states on the analogy of awareness of pressure ("touch") which "involves a direct, or non-inferential, awareness of causes,"11 fails. For awareness of mental states (dispositions) must, since dispositions ≠ physical items simpliciter, be inferential: to know I'm in pain, I must know that that of which I am aware meets further conditions relevant to realizing a functional syndrome—I must infer that I'm in pain.

It is thus difficult to see what non-eliminativist account can be given the objects of introspection. We can avoid the eliminative claim by saying that when the disposition is manifested—when the functional type is realized—felt pain supervenes. But we have seen the difficulties with this tack. Furthermore, to claim that supervenience is an
explanatory thesis, not an ontological one, appears elimina-
vivist in that felt pains are, again, constructed entities
with a heuristic value but lacking true ontological status.
(Wimsatt in "Reductionism, Levels of Organization, and the
Mind-Body Problem,"\textsuperscript{12} appears to argue for something like
this, though I am not sure.)

It may be claimed that there is some question begging
going on in that the data of introspection are being used
to deny (functionalist) token-identity, but if the token-
identity claim is correct, introspection is but one part of
the brain scanning another (or one part of the whatever
scanning another)—a purely physical process employing only
physical items. And if being aware of something as an
object of experience is the same as the what it is like of
experience, then the latter is a physical feature. Thus
the token-identity theorist may demand an independent argu-
ment for the claim that the what it is like feature of pain
cannot be accounted for on his view. I cannot give an
argument that is independent of the data of introspection—
it is basic—but I shall argue that the token-identity
theorist cannot provide an account in physical/functional
terms which will fill the role that experience provides in
our account of the world. It will be argued that some con-
cepts, for instance, the pre-scientific concept of pain,
with which any theory of pain need reckon, must be ostensively
learned or identified via the qualities of experience, the hurtings, that figure as the objects of experience and which are recognized, named, and discriminated by experiencers (perhaps I should say, "community of experiencers"). This leads directly to the Private Language Argument (assuming there is one such), for it is often thought that this argument shows it meaningless to speak of objects of experience (to countenance their existence) where such objects are in some sense private to the experiencer. It is argued that there is a sense in which these objects are private; we accept the sincere reports of accomplished (sensation) language users as the best (in some cases conceivably the only) evidence, the best (or only) criterion for the correctness of an individual's claim to be experiencing a sensation of a particular sort. It is not argued that there is, or could be, a purely private or ostensive language; nor need it be argued.

Private Language\textsuperscript{13}

Recently Saul Kripke has reviewed interest in Wittgenstein by reconstructing the latter's skeptical position that citation of a rule is insufficient to warrant any assertion that what we have done on a particular occasion is, or is not, what is specified by the rule cited. In his discussion, Kripke argues that Wittgenstein sets out a skeptical paradox which entails the conclusion that "[t]here can be no such
thing as meaning anything by any word."\(^{14}\) As Wittgenstein says in \$201, "This was our paradox: no course of action could be determined by a rule, because every course of action can be made out to accord with the rule."\(^{15}\) It is with regard to Wittgenstein's attempted solution to the paradox, maintains Kripke, that the argument against private language appears.

After developing the skeptical paradox, I shall argue, following hints by Kripke, that talk of inner states is legitimate and that a liberalized version of the private language argument is consistent with Wittgenstein's solution of the paradox and the result that some rules cannot be checked by others once the rule-follower has been admitted as an accomplished language user. It is concluded that first-person reports of mental states provide sufficient grounds for justifying the claim that there are "private" psychological (phenomenal) items of which, say, pain feelers are aware and to which they make reference via their reports (labeled by Kripke "avowals"). "Pain," for instance, is made the name of +pain+ in that it stands for whatever has certain natural expressions (typical pain-behavior)—it is contingent that +pains+ have such natural expressions. This is not to say "pain" is defined as whatever has certain natural expressions; +pain+, rather, is that particular item to which "pain" refers even though
the meaning (=use) of "pain" depends on public stage-setting (see below).

As Kripke suggests, a primitive feature of the language game (I leave this undefined) is that once a speaker is judged to have mastered (say) sensation language, the speaker's sincere report of pain is taken as the criterion for its correctness. Evolution, I take it, justifies my belief (though not with certainty) that when others sharing similar genetic endowment, sincerely report their pains, they feel something like that which I feel when I sincerely report my pains, viz., +pain+. Moreover, this qualitative feature cannot be fully understood merely by knowing all the physiological or functional facts about pain feelers. Nor can X form an appropriate attitude toward Y's being in pain if X has never felt pain. Although public circumstances (including verbal reports) may ground your attribution of pain to me, and vice versa, as well as enable us to learn the use of "pain", it would be a mistake to hold that all there is to felt pain, all the facts there are, are publicly ascertainable via objective science. What is obviously correct in Wittgenstein's charge is that the private accompaniments of pain-behavior cannot be sundered from their (in principle) public manifestations if we are to investigate felt pains—obviously, in that only via (directly or indirectly, see Chapter 6) reports of experiencers of pain
can we be warranted in claiming that another feels pain. Mere scientific description in terms of physical/functional facts cannot shed adequate ontological understanding on +pains+, nor can science ignore them. Let us turn to Kripke and Wittgenstein.

Kripke sets the paradox with a mathematical example. It is typically thought that when we solve an addition problem, e.g., $68 + 57$, we are successful because we "grasp" the rule for addition; where such "grasping" involves some type of internal mental representation. Although the rule applies to indefinitely many cases of addition, each of us has actually computed only finitely many sums in learning the rule. This starts the problem. Assuming that "$68 + 57$" is an example I have never computed, a skeptic challenges the claim that the answer is 125. He claims that in the past I have never computed numbers larger than 57, so that now I should answer 5. Why? Because, he claims, it is possible that in the past I used "plus" and "+" to denote a function called "quus", symbolized by "©" and defined thus:

\[ x \circ y = x + y, \text{ if } x, y < 57 \]
\[ = 5 \text{ otherwise.} \]

In effect, the skeptic's challenge is that I am now misinterpreting my past usage (rule following) since I always meant quus by "plus". As Kripke remarks about the
skeptic's hypothesis, "[w]hile it indubitably is, no doubt it is false; but if it is false; there must be some fact about my past usage that can be cited to refute it." A typical response to the skeptic would be that 125, not 5 is the answer because it accords with what I meant by "+". Kripke points out, however, that

[an answer to the sceptic must satisfy two conditions. First, it must give an account of what fact it is (about my mental state) that constitutes my meaning plus, not quus. But further, there is a condition that any putative candidate for such a fact must satisfy. It must, in some sense, show how I am justified in giving the answer "125" to "68 + 57".]

As we shall see, the challenge is not an epistemological one. Any answer to the skeptic, then, must present a candidate for a fact as to what I meant that would show that only "125", not "5", is the answer I ought to give. It seems natural to respond that in learning addition one learns a rule which determines how addition (of new problems) is to be continued—the concept of addition is not learned, per the impossible, by memory of specific examples. By hypothesis I have never computed for arguments over 57. Again, consider Kripke's clear posing of the problem:

What was the rule? Well, say, to take it in its most primitive form: suppose we wish to add \( x \) and \( y \). Take a huge bunch of marbles. First count out \( x \) marbles in one heap. Then count out \( y \) marbles in
another. Put the two heaps together and count out the number of marbles in the union thus formed. The result is \(x + y\). This set of directions, I may suppose, I explicitly gave myself at some earlier time.19

But with regard to this learned algorithm for addition, Kripke gives the skeptic’s reply:

True, if "count", as I used the word in the past, referred to the act of counting. . . , then "plus" must have stood for addition. . . . he [the skeptic] can claim that by "count" I formerly meant quount, where to "quount" a heap is to count it in the ordinary sense, unless the heap was formed as the union of two heaps, one of which has 57 or more items, in which case one must automatically give the answer '5'.20

Quickly we run into Wittgenstein's comments about "a rule for interpreting a rule." The skeptical strategy can be reiterated for any rule, including any "basic" rule; any present application of such a rule is not justified and the rule is applied blindly. So my response to the problem "68 + 57" can be "5" or "125"—I have no internal justification for the latter as opposed to the former since there is no fact about me which distinguishes between my meaning plus rather than quus.

Although I am ignoring many of the niceties that Kripke brings out, I take it that the skeptical problem begins to take shape—in full force, Wittgenstein's problem is that
he has apparently "shown all language, all concept formation, to be impossible, indeed unintelligible." Consider again the view that there is a meaning I attach to the "plus" sign which instructs me in future computations. Kripke reviews:

But when I concentrate on what is now in my mind, what instructions can be found there? How can I be said to be acting on the basis of these instructions when I act in the future? The infinitely many cases of the table are not in my mind for my future self to consult. To say that there is a general rule in my mind that tells me how to add in the future is only to throw the problem back on to other rules that also seem to be given only in terms of finitely many cases. What can there be in my mind that I make use of when I act in the future? It seems that the entire idea of meaning vanishes into thin air.

The link between the mathematical example and the language of sensations can easily be established (leaning on Nelson Goodman. It would seem that when we determine our use of the word "pain", we only need to recall--bring to mind--what a particular paradigm felt pain is like in order to apply the word in the future. But this justification of my application of "pain" to a new sensation leads to the skeptical problem. We can imagine the skeptic saying that by "pain" in the past I meant pickle (where "pickle" applies to pains before $t$, and tickles thereafter) and the sensation which was pickle was meant to direct me to apply the word "pain" to pickle sensations always. If the tickle sensation
I am aware of now is pickle, then it falls in the extension of "pain" as I meant it in the past. Nor can the difficulty be avoided by stipulating that "pain" applies to all and only those sensations with the same feel as my paradigm. Our skeptic can reinterpret "same feel" as same "schmeel", where things have the same schmeel if . . .

For present purposes I shall take it that the skeptical paradox is truly a paradox; that is, I will agree with Kripke's skeptic "that there is no "superlative" fact (§192) about my mind that constitutes my meaning addition by "plus" and determines in advance what I should do to accord with this meaning." (Kripke does argue that a dispositional analysis of the fact that I meant plus (rather than quus) is not successful in avoiding the paradox: "Precisely the fact that our answer to the question of which function I meant is justificatory of my present response is ignored in the dispositional account and leads to all its difficulties. Wittgenstein too, argues Kripke, accepts the paradox, but offers a skeptical solution--a solution that concedes that the skeptic's claims cannot be answered in a certain way. The justification the skeptic shows to be untenable is not, if Wittgenstein is correct, required for language use and concept formation.

The conclusion drawn from the skeptical paradox is that "plus" ("pain") does not have a meaning--there is no fact as
to whether I mean plus or quus—in terms of truth-conditions. Indeed, in full force there is no fact about what I mean by any word at any time and no justification in terms of facts which supports the claim that I should answer "125", not "5", to "68 + 57"; hence, the claim that I mean plus by "plus" is meaningless. Before turning to Wittgenstein's solution, let us examine James Cornman's understanding of the Private Language Argument, for my response can more clearly be seen with his aid.

Cornman is concerned with the claim that the Private Language Argument can be used to conclude that first-person psychological sentences are not reports of private psychological entities. He reconstructs the argument as follows:

(1) If at most one person, A, can know any rules for the use of 'E', then no one can distinguish the difference between someone obeying and merely thinking that he is obeying the rules.

(2) If no one can distinguish the difference between someone obeying and merely thinking that he is obeying a rule, then no one can know whether someone is obeying the rule.

(3) If no one can know whether someone is obeying a rule, then no one can obey the rule.

(4) If no one can obey any rules for the use of an expression, then no one can know how to use the expression correctly.

(5) If no one can know how to use an expression correctly, then the expression is meaningless.

Therefore
(6) If at most one person, A, can know the rules for the use of "E", then 'E' is meaningless.26

We see that this first part of the argument is directly tied to what has gone above. Things are a bit different, of course, for Cornman does not have the skeptical paradox in mind. Given the skeptical paradox, it follows that no one can distinguish the difference between someone obeying and merely thinking that he is obeying the rules for, e.g., "plus". Given that there is no fact about me (my mind) which distinguishes between my meaning plus rather than quus, I have no (internal) justification that I am correctly applying the rules for "plus". Cornman argues that the only doubtful premise is (1).27 He argues that memory can serve to justify A in thinking he is correctly applying the rules for "E". As Kripke would point out, this ignores the skeptical challenge:

Given, however, that everything in my mental history is compatible both with the conclusion that I meant plus and with the conclusion that I meant quus, it is clear that the sceptical challenge is not really an epistemological one. It purports to show that nothing in my mental history of past behavior--not even what an omniscient God would know--could establish whether I meant plus or quus, . . . There can be no fact as to what I mean by 'plus', or any other word at any time. The ladder must finally be kicked away.28
Let us not linger over (1), but grant that memory will not serve to justify A in determining whether he correctly or incorrectly thinks he is obeying the rules. As we shall see, Wittgenstein's tack is to deny the antecedent of (1) and inveigh against the private model of rule following. In the second half of his version of the argument, Cornman centers concern on alleged private entities. After giving Cornman's version, I shall turn to Kripke and Wittgenstein for the details. Cornman reconstructs as follows:

(7) If E is a private entity of A, then at most A can know whether it exists or occurs.

(8) If at most A can know whether E exists or occurs and 'E' is used to denote E, then at most A can know whether 'E' actually does denote E.

(9) If at most A can know whether 'E' denotes E, then at most A can know the rules for the use of 'E'.

Therefore

(10) If E is a private entity of A and 'E' is used to denote E, then 'E' is meaningless.

(11) No psychological expressions such as 'sensation', 'pain', and 'afterimage', are meaningless.

Therefore

(12) No psychological expressions are used to denote private psychological entities.

Cornman considers (9) to present the most crucial problem for the argument; I agree. To flush out the difficulty
with (9), however, we must first turn to Wittgenstein's attempted solution to the paradox, for it is the solution which disallows private language.

For present purposes I shall limn (à la Kripke) Wittgenstein's solution to his paradox, pointing out the sense in which the solution will not allow a private language. I have often heard the claim that for Wittgenstein "meaning is use"; as I understand it the aphorism, "Don't think, look!", points to the notion that use determines meaning. By replacing talk of truth-conditions with talk of conditions of assertability, and then seeking the utility of making assertions under specified conditions within our linguistic framework, Wittgenstein seeks to legitimize meaning claims (e.g., "Jones means plus by "+" ") by holding that the rules for correct usage are public rules employed by a community of language users. Kripke renders this notion thus:

All that is needed to legitimize assertions that someone means something is that there be roughly specifiable circumstances under which they are legitimately assertable, and that the game of asserting them under such conditions has a role in our lives. No supposition that "facts correspond" to those assertions is needed.30

This is in stark contrast to the earlier notion, eschewed because of the paradox, that a person is correctly following the rule for addition because he has "grasped" such a rule and hence, communication and concept formation
proceed because each of us (privately) grasps the same concept. Thus the preceding argument can be seen as a reductio on the antecedent of premise (1). If we consider an individual in isolation (not as a member of a language community), and investigate his internal psychological state and his behavior, we will be unable to find a fact about him that will warrant the claim that he should act according to the plus function rather than an alterantive. It makes no sense to claim he is correctly or incorrectly following a rule—he is justified in applying the rule in any way that strikes him. But if meaningful language is a rule governed activity, this model (of private rule following) forces us to the view that meaningful language use is chimerical. But obviously it is not. Wittgenstein's solution is to avoid this model of private rule following. In §202 he states:

And to think one is obeying a rule is not to obey a rule. Hence it is not possible to obey a rule "privately": otherwise thinking one was obeying a rule would be the same thing as obeying it. 32

Jones' thinking he is obeying the rule for addition because he believes he has grasped the concept of addition, does not justify his claim that "125" is the answer to "68 + 57" since the answer "5" would be as justified. Consider the conditional, "If Jones means addition by "+", then if he is
asked for "68 + 57", he will reply "125". Instead of understanding this to imply that a mental state of Jones guarantees that he means addition, Wittgenstein would focus on the inversion (barring computational error, etc.), "If Jones does not come out with "125" when asked about "60 + 57", we cannot assert that he means addition by "+"."33 What we end up with is the view that the rule for addition is a conventional rule where its correctly being followed is a function of its checkability by (in terms of agreement with) the community--rule following and meaning (=use) is a public affair.

So, for Wittgenstein, the function of language is not to state facts. What, then, is its function? One might say, only half jokingly, it is a many splendored thing; for to think of just the role played by the practice of attributing sensation concepts, when someone sincerely utters, "I am in pain," that utterance will (typically) invoke in others an attitude which will (typically) induce them to offer comfort, say, an analgesic. This is very close to remarks by Ayer, for it would seem that sensation terms, like Ayer's ethical terms, "are calculated also to arouse feeling, and so to stimulate action."34 Be this as it may, this is not the only function such utterances serve. It may be thought that felt asymmetry between first-person and third-person utterances vanishes because so called first-person reports are not
(really) reports. That is, it may be tempting to conclude that "pain", for instance, does not refer to +pain+ since the conditions apt for its assertability are its natural expressions—typical pain-behavior and concomitant external circumstances. (This seems to be in accord with premise (12) above.) There is some evidence in Wittgenstein's writings that he considers expressions such as "I am in pain" as expressive. But there is also evidence that he did not consider them always as merely expressive; nor should he, for when I say to the doctor, "I have an intense, shooting pain in my lower back", I am informing him of the type of pain, its location, intensity, etc.—I am describing my inner state and this is not analogous to a cry "escaping" from me.

In §580, Wittgenstein says "An 'inner process' stands in need of outward criteria." But this is to be understood in the context of attributing to others mastery of the language of sensations, and figures prominently in the solution of the paradox. Consider Kripke:

Roughly speaking, outward criteria for an inner process are circumstances, observable in the behavior of an individual, which, when present, would lead others to agree with his avowals. . . . it is part of Wittgenstein's general view of the workings of all our expressions attributing concepts that others can confirm whether a subject's responses agree with their own.
A sceptical problem is posed, and a sceptical solution to that problem is given. The solution turns on the idea that each person who claims to be following a rule can be checked by others. Others in the community can check whether the putative rule follower is or is not giving particular responses that they endorse, that agree with their own. The way they check this is, in general, a primitive part of the language game.  

Thus, not only are outward criteria necessary for learning (teaching) the use of words, but also for checking that another uses them correctly (as we do). It should not be concluded, however, that a person is in pain if the outward criteria are satisfied or is not in pain if the criteria are not satisfied—where, that is, said criteria may conflict with the person's sincere avowal. As Kripke conjectures, it seems empirically false that inner processes always have outward criteria, for it may well be that there are some "sensations or sensation qualia that we can perfectly well identify but that have no "natural" external manifestations; an observer cannot tell in any way whether an individual has them unless that individual avows them."  

This allows for a liberalized version of the private language argument, for the language of the sincere speaker,

even his language of sensations, will not have the objectionable form of a "private language", one in which anything he calls "right" is right. The speaker can demonstrate, for many sensations that do have "public criteria", that he has mastered
the appropriate terminology for identifying these sensations. If we agree with his responses in enough cases of various sensations, we say of him that he has mastered "sensation language". . . . if an individual has satisfied criteria for a mastery of sensation language in general, we then respect his claim to have identified a new type of sensation even if the sensation is correlated with nothing publicly observable. Then the only "public criterion" for such an avowal will be the sincere avowal itself.39

Compare this version with the case of Bat, who we imagined is the only person capable of feeling pain. It is not as though Bat, or anyone, in a purely private way thought, "This is what I will call "pain" from now on." In mastering sensation language (a community affair) lots of public stage-setting40 was necessary for Bat to come to single out some of his inner goings-on, namely +pains+, and refer to them via "pain". In Bat's case, the stage-setting may even come about by way of written accounts of past pain feelers. Others may know the conditions when it would be appropriate to assert of Bat that he is in pain--in fact, Bat can teach them--though they will not know what Bat knows, viz., what it is like to feel pain. Consider how a normal child is taught the use of the word "pain". His "teachers" clue him when it is appropriate to say, "I am in pain," "She is in pain," etc., via public circumstances (e.g., upon the occasion of a bee stinging Betty on the arm)--these circumstances amount to the "stage-setting".
The child's use of pain-language can be checked by the community; given that "pains" are what accompanies the "natural" expressions which serve to set the stage for the use of "pain", this stage-setting enables the child to sort out his inner goings-on and evolution (shared genetic endowment) justifies the view that the child singles out "pain" as what accompanies his pain reports and as what enables him to form the "appropriate" attitude toward others (e.g., Betty) who report pain or are attributed pain (more on this below).

It is, of course, logically possible that the child singles out "fear" rather than "pain", but, in the long run, not probable. Nor, and this I take it is Wittgenstein's point, does it matter, in order to assure uniform language use, what is picked out as the accompaniment of such reports as long as the child "goes on" in the same way we do. Let me develop the claim that Wittgenstein does not (or should not) deny that "pain" refers to (names) that which accompanies the natural expression which gives meaning--allows uniform and checkable use or fule following--to "pain".

As a starter I again turn to Kripke, but shall develop the distinction between Bild (picture) and Vorstellung (image) a bit more. Kripke tells us:

I am inclined not to accept the conclusion I have sometimes heard drawn that for Wittgenstein my inner experience of pain,
and my ability to imagine the sensation, play no real role in my mastery of the "language game" of attributing sensations to others, that someone who has never experienced pain and cannot imagine it but has learned the usual behavioral criteria for its attribution uses this terminology just as well as I.41

(Of course I would add to this citing of behavioral criteria, physiological/functional criteria.) When we attribute +pain+ to another do we not, on Wittgenstein's view, use behavioral criteria? We may see a bee sting Betty and hear her cry out, or we may witness others rushing to her aid (the frantic mother). Cannot a person unable to feel pain use this criteria as well as we? Consider §297:

Of course, if water boils in a pot, steam comes out of the pot and also pictured steam comes out of the pictured pot. But what if one insisted on saying that there must also be something boiling in the picture of the pot?42

If we think of what we witness when we see a bee sting Betty and hear her subsequent cries as her mother rushes frantically to her aid, on analogy with what we view upon looking at a picture, then +pain+ (cf. boiling liquid in the pictured pot) itself is not pictured, only pain-behavior. Surely, this we can grant. But our interpretation and description of such a "picture" does depend on our assumptions about what backs the pain-behavior--think of a scene in a play. We may, at a play, witness an actor exhibit convincing
pain-behavior, yet we do not rush to his aid, though we may feel the urge to do so; that is, we do not conclude he truly experiences +pain+. Nor can we describe the picture of the pot as a picture of steam escaping from the pot without reference to something in the pot. Depending on what we assume to be in the pot (as well as contextual clues), we may think it a picture of Aladdin's lamp. How is the picture made complete? Wittgenstein makes a distinction between Bild and Vorstellung that helps answer this question. In §300, we find:

It is—we should like to say—not merely the picture of the behaviour that plays a part in the language-game with the words "he is in pain", but also the picture of the pain. Or, not merely the paradigm of the behaviour, but also that of the pain.-- It is a misunderstanding to say "The picture of pain enters into the language-game with the word "pain"." The image of pain is not a picture and this image is not replaceable in the language-game by anything that we should call a picture.-- The image of pain certainly enters into the language-game in a sense; only not as a picture.43

The picture of pain does not enter into the language game with the word "pain" because, although we can imagine pain (+pain+), we cannot (in the case of another) picture it (though we can feel it in our own case). All we witness, as in a play, are its natural expressions, etc. Still, the image of pain enters into the language game in that we can
imagine the pain we would feel in such circumstances—thus cathexis as we watch a play and feel the urge to comfort or avenge. When we predicate "pain" of (the non-actor) Jones, do we make reference to +pain+? Indeed we do, for to distinguish between the pain behavior of the actor-Jones and the "same" behavior of the non-acting Jones, we imagine the existence of +pain+ in the latter but not the former. Although, if the acting is done well enough, we typically feel sympathy in both instances (although only if we ourselves are amongst the class of pain feelers), only the latter typically excites us to action. Why? Because, as opposed to mere behavior, a reference to +pain+ is made in the latter.

For the person, Smith, congenitally unable to feel pain, it strikes me that the quality of Smith's attitude toward Jones (under these two guises) will differ from the attitude of pain feelers; though Smith's behavior may not differ. (Kripke suggests that use of the image as a picture corresponds to holding that there are determinate truth-conditions for another's being in pain; a move, he claims, Wittgenstein would deny.) It certainly seems to me that I do have an attitude (of sympathy) toward another in pain that I would not have were I unable to feel pain, unable to imagine that I experienced the +pain+ I attribute. With regard to how the image of pain enters the language game,
I agree with Kripke:

My suggestion is that it enters into the formation and quality of my attitude toward the sufferer. I, who have myself experienced pain and can imagine it, can imaginatively put myself in place of the sufferer; and my ability to do this gives my attitude a quality that it would lack if I had merely learned a set of rules as to when to attribute pain to others and how to help them.44

As I put this point, an experiencer of +pain+ can know what it is like, thus forming an appropriate attitude, whereas a person unable to feel pain cannot know what it is like. It is the role played by this what it is like feature that cannot be played by mere physical/functional items.45

"Once again, the correct interpretation of our normal discourse involves a certain inversion: we do not pity others because we attribute pain to them, we attribute pain to others because we pity them."46

I take it, then, that "pain" does not refer to that which has its natural expression in pain-behavior—a private something. Nonetheless, I am not committed to a (purely) private language, for the meaning of "pain" is not dependent on this something. That only Jones can have final authority as to whether or not he is in pain (experiences +pain+) does not entail that only he can know the rules governing the use of "pain". By making this plausible, I make it plausible that the existence of "private" objects does not entail the
objectionable form of a private language, namely, private rule following. Thus, premise

\[(9) \text{ If at most } A \text{ can know whether 'E' denotes } E, \text{ then at most } A \text{ can know the rules for the use of 'E',}\]

is false. And this follows even if we grant the truth of (1). Further support for the falsity of (9) comes from considering §304:

"But you will surely admit that there is a difference between pain-behaviour accompanied by pain and pain-behaviour without any pain?"--Admit it? What greater difference could there be?--"And yet you again and again reach the conclusion that the sensation itself is a nothing."--Not at all. It is not a something, but not a nothing either! The conclusion was only that a nothing would serve just as well as a something about which nothing could be said. We have only rejected the grammar which tries to force itself on us here.\]  

What can be said about +pain+ (in physical/functional terms) that would enable one not able to feel it to understand that to which "pain" referred? A person unable to feel pain would never know to what "toothache" referred: "a something about which nothing could be said." Could such a person understand the meaning (=use) of "toothache"? Turning to Wittgenstein we find:
"What would it be like if human beings shewed no outward sign of pain (did not groan, grimace, etc.)? Then it would be impossible to teach a child the use of the word "toothache."

And when we speak of someone's having given a name to pain, what is presupposed is the existence of the grammar of the word "pain"; it shews the post where the new word is stationed. 48

Since human beings typically do show outward signs of pain, our imagined person could, in this sense, come to use the word "toothache", though, per the above, he would not know that to which it referred. So if he mastered (the use of) sensation language, he may understand that "toothache" refers, but he need not know to what it refers for the purpose of learning the meaning of "toothache": "a nothing would serve just as well." Even those (most) of us who do feel pain, use "pain" to refer to that which has its natural expression in pain-behavior only because of public stage-setting. But it would be wrong to conclude that "toothache" has no place in the language game at all. Wittgenstein appears to so conclude, for in §293 we find:

Now someone tells me that he knows what pain is only from his own case!—Suppose everyone had a box with something in it: we call it a "beetle". No one can look into anyone else's box, and everyone says he knows what a beetle is only by looking at his beetle.
The thing in the box has no place in the language-game at all; not even as a something: for the box might even be empty. — No, one can "divide through" by the thing in the box; it cancels out, whatever it is.49

Thus, if the word "toothache" referred to what each of us experiences, then the meaning of the term "toothache" (cf. "beetle") would not depend on what each experienced, if anything. Given this passage, Wittgenstein seems to conclude that what "toothache" refers to, or whether it refers at all, is irrelevant to the language game (cf. §274). As we have seen, however, the existence of the "object" in the box— that which accompanies pain-behavior— does play a role in the formation of our attitudes toward fellow pain feelers and enables us to distinguish between "real" life pains and pains-in-plays. It is, of course, possible that what accompanies my pain-behavior may differ radically from what accompanies yours, but evolution and commonsense justify us in thinking otherwise.50

Premise (9) implies that a necessary condition of being able to grasp the rules for the correct use of an expression is being able to know the reference of the expression. But if we can come to understand the use of some particular expression independently of knowing its reference . . . then (9) is surely false.51

Consider an objection raised by Lycan (in correspondence);
Suppose Jones and Smith have both mastered the assertibility-conditions for pain avowals, and on one occasion both make pain avowals under (publicly) appropriate circumstances. Yet if Smith is (despite this) not actually feeling pain, his utterance is false while Jones' avowal is true, and so they must mean something different even though their "assertibility-meanings" are identical.

As we have seen, Wittgenstein can be understood as claiming that use determines meaning in the sense that knowing the use of an expression, knowing its assertibility-conditions, suffices for knowing the meaning of the expression. Lycan's (counter-) example gives voice to the intuition that meaning determines truth-conditions; but if this is so, then it would follow that use determines truth-conditions. Yet Wittgenstein inveighs against truth-conditions; thus he must deny that meaning determines truth-conditions. But Lycan's example strongly suggests that meaning does determine truth-conditions. I think Lycan's example can be used in my favor, as a critical summation of Wittgenstein's argument makes clear.

Consider some brief comments by Feigl:

Our world, being what it is, can, of course, be known by description in any of its parts or aspects, but only on the basis of a foothold somewhere in acquaintance.52

In the pursuit of a logico-epistemological reconstruction of our knowledge claims regarding the external or physical world
we are— in the last analysis— driven back to the phenomenally given as the ultimate testing ground.53

In learning a language we cannot expect success via only verbal explanations of meanings, that is, not all knowledge of meaning can be explicit verifiable knowledge. As I interpret (and stretch to fit the present discussion) Feigl, at some point a learner's grasp of meaning must consist in something other than an awareness that one sentence means the same as another. Let us concede, then, that the language-learner's ability to grasp meanings must somewhere involve confrontation with the items to which words refer.

Granted this point, we can cast much of the controversy centering around the notion of a private language as being generated due to the fact that the language-learner's attachment of meaning to an expression of (sensation) language will amount to a construction out of what he has experienced. For the language-learner must select from experience what his teacher intends to ostend; but the traps here are obvious, as Wittgenstein points out. Yet since we do successfully learn language and engage in meaningful communication with one another, the traps are typically avoided— we do learn, via ostension, that to which, e.g., "ball", "sock", "shoes", etc., refer. Disagreement settles in when we turn to the question of the relevance (as far as
meaning is concerned) of what is and is not experienced. Is the meaningfulness of "pain" tied to (grounded in) a "private object" or to some publicly ascertainable items?

Wittgenstein appears to argue that the language-learner can learn the meaning of a term only if its meaning is tied to publicly observable features. This is the "stage-setting", and unless the teacher observes (in the appropriate circumstances) pain-stimuli and/or subsequent pain-behavior ("natural expressions"), learning cannot begin. Suppose we concede this; how do we move from this to the claim that use determines meaning?

Should we deny that a speaker's grasp of meaning must consist in something other than an awareness that one sentence means the same as another, denying, that is, that there must be implicit knowledge, then an infinite regress follows—in order to acquire linguistic abilities one would already have to have linguistic abilities (cf. "a rule for interpreting a rule"). (Of course we are to construe these "linguistic abilities" as of a semiotic nature and not purely syntactic, since the latter may well be innate.) When, for example, our language-learner finally says "I know what "pain" means", we will agree only if she understands her words. And we will agree to this (goes the argument) only if her understanding the meaning of her words (her correct application of rules) can be checked
otherwise than by her being able to say what she does (thinking she is obeying a rule). Her understanding, her speaking meaningfully (by our lights) is granted only if her implicit knowledge is manifested by way of an observable difference between her behavior and the behavior of one (e.g., an idiot, a parrot) who does not have such knowledge—e.g., mere utterance of vocables does not constitute meaningful language use. Her understanding must issue in some practical ability, not merely the ability to state that such-and-such is the case.

Thus, one may wish to claim that the ascription of meaningfulness to (say) utterances is backed by the publicly observable aspects of behavior manifested by use of language in appropriate circumstances. This in turn, it seems, is thought necessary because if awareness of something private was thought essential to my speaking meaningfully when I say, for example, "I am in pain", then others could not know I knew (had an understanding of) what I said (that I was consistently and correctly applying rules) unless they knew of the existence of the private object. But, as we have seen, you cannot know of (be aware of) my private object (although, as Castañeda maintains, this need not be a logical impossibility).

So, to ascribe meaningfulness to one another's utterances, we must know that we apply rules consistently, we
must know that one another understands the words used—meaningful communication is a community affair. But if all we can know is what is publicly observable, private objects are not essential to meaning—that is, if we are to avoid the absurdity that language is meaningless. Hence it is concluded that meaning determines assertability-conditions having to do with publicly observable circumstances and behaviors. I do not think Wittgenstein claims that an individual does not associate with a word some mental content not lying in the use he makes of the word, but that such mental content would be ineffable and irrelevant to meaning—"a something about which nothing could be said."55

But we are able to communicate, sensation language is meaningful and hence, it is concluded, sensation language (first-person avowals) must be grounded in (in principle) publicly observable conditions. If it were not, language-learning could never take place, it would have no foothold in private (objects) of experience. (But, of course, why think these "publicly" observable conditions must be assertability-conditions?) Compare Wittgenstein:

The individual words of this [private] language are to refer to what can only be known to the person speaking; to his immediate private sensations. So another person cannot understand the language.57
Here it is suggested that if there were anything private to, e.g., Jones which was essential to the meanings of his words, those meanings would be incommunicable. Thus Jones' avowals could not be understood (because we cannot know that to which he refers?). Let us return to the language-learning situation.

It is true that the teacher, T, must assume that the student, S, is aware of (say) +pain+ (in the appropriate circumstances), if private ostensive definition is to serve in explaining (sensation) language-learning (cf. chapter 2 above). Let us also concede that T can never know (be certain!) that S is aware of +pain+ (or correctly attaches "pain" to +pain+); still it does not follow that T cannot understand S's (nor S understand other's) avowals just because T (S) cannot know that to which S (others) refer (this is the "selection problem"). That is, it strikes me that all we need for successful communication and sensation language use (learning) are justified true beliefs that another who avows she is in pain has experiences like ours—we need no guarantee, we do not need to know this assumption true. Does our inability to know this assumption true entail that it is not rationally justified? That is, why think rational justification (that you mean by "pain" what I mean) depends on our knowing (as adumbrated) what one another means and thus, on publicly observable conditions—even if
meaning is a community affair? If this is not an objectionable form of verificationism, it is surely close.

Maybe (it certainly seems true that) I cannot be aware of (know by acquaintance) your private inner mental contents, but it does seem (especially given evolutionary considerations) that felt pains are a kind of thing of which I can be (and am) aware. This, I maintain, is all that is needed for language-learning and successful communication (meaningful language use). Feigl hints at something close to this:

From the point of view of the liberalized criterion of factual meaningfulness we don't have to restrict the meaning of mental state ascriptions to the behavioral "criteria." The criteria of which the neo-Wittgensteinians speak are allegedly quite different from symptoms. Nor are they to be understood as logically equivalent to, or entailing, the ascriptions in question. They can serve as (empirically?) necessary and sufficient conditions of those ascriptions only under "normal" circumstances. Perhaps this can be accepted as a fairly adequate analysis of the way we actually ascribe--in ordinary situations and in terms of the commonly used language--mental states to other persons. But is there then really that essential difference between (fairly reliable) symptoms and criteria?58

This is not to say that a purely private language is (or is not) possible, only that the sense that attaches to expressions does transcend "use". Our ability to use language to communicate with and understand one another can
be accounted for by allowing private ostension a (primitive) place in the "language game" (cf. the liberalized version of PLA, pp. 203-04 above). Thus we can allow private ostensive definition (reference to private objects) without language becoming meaningless. Indeed, the very plausible assumption that we do experience sensations of the same kind serves to justify the claim that we mean the same thing by "pain"—and this even though we do not (or cannot come to) know the assumption true.

Now, what Lycan's objection points out is that our attributions of understanding on the part of a speaker, which are to ground his expressions as meaningful, cannot themselves be grounded in public assertability-conditions. Intuitively, the claim that assertability-conditions can be met and yet Jones mean something different from what Smith means is quite convincing. Thus, there is plausibility to the notion that there is a fact about Jones which transcends use and which is relevant (i.e., essential) to what his words mean. What Jones (Smith) means depends not merely on assertability-conditions for pain-avowals, but also (to put it vaguely) speaker reference, where such reference may be to private items not captured by publicly observable circumstances and behaviors.

Again, compare Feigl:
The arguments of the Wittgensteinians . . . are utterly implausible to me. Introspective reports or avowals are either true or false. . . . Reference to one's own immediate experience is the (epistemological!) prototype of all designations of objects, properties or relations by the words of our language.59

One may attempt to run another "private language objection" in terms of truth-conditions by starting from the claim that the fact in virtue of which we know Jones is in pain (his pain-behavior) is a different fact from the fact in virtue of which it is true that Jones is in pain. The next move would be to claim that such facts must be publicly ascertainable if we are ever to give a completely satisfactory account of meaning. Such a program would simply replace Wittgenstein's assertability-conditions by truth-conditions. I cannot pursue such a program here, but it is clear that the threat of verificationism hovers close by—why think the facts grounding truth-conditions must be publicly ascertainable? Pursuit of this program, it seems to me, would once more lead us to the claim that that which is grasped in experience cannot give us knowledge (other than topic-neutral knowledge) of the facts that make our first-person reports true (or false). If this (admittedly sketchy) notion concerning truth-conditions is correct, we would simply be off and running over the issues discussed in previous chapters. Suppose "Jones is in pain" is true if
and only if Jones realizes functional syndrome $F$ that is of the $S_p$ kind. How will this provide the non-pain feeler with knowledge of what "pain" is, of what "pain" means?

Consider (what may possibly be) another way of making the same point. Herbert Heidelberger concludes:

A person who understands a sentence—one who is aware of what the sentence expresses—is a person who is aware of the conditions under which the sentence is true; but a person may be aware of the conditions under which a sentence is true without understanding the sentence. . . . I suppose that we could express this moral more simply—although somewhat cryptically—by saying that understanding a sentence is more basic than being aware of the sentence's truth conditions. 60

Suppose (the non-pain feeler) Smith (truly) believes: ["Jones is in pain" is true if and only if Jones is in physiological state $n$ that is of the $S_p$ kind]. It does not follow that Smith understands "Jones is in pain". Without himself being (or having been) aware (de re) of the kind of item Jones is aware of when in the pain state, Smith fails to understand the sentence, even though he is aware of the sentence's truth-conditions. As I put this point, Smith is only aware (de dicto) of that which causes Jones to experience pain; he is not aware of what Jones expresses when he utters "I am in pain", or otherwise affirms that he is in pain). I use "expresses" as a primitive to label what intuition leads me to believe true--Jones is communicating
something about his private experience that we understand only if we too have had like kinds of experiences. That I (or Jones) cannot say what fact about myself (him) is expressed does not mean a fact is not expressed. You, as a pain feeler, know what fact about me is expressed when I report or otherwise indicate that I am in pain— it is what happens to me when I realize a functional state $S_p$: I become aware de re of +pain+, a phenomenal (nonphysical) property instance that is the private (quality-) object of my awareness. You understand because you too (sometimes) have a like awareness. Smith does not understand even though he is aware of the conditions under which it is true that I (you) become aware of +pain+. (The materialist denies this: are we at the point of "conflicting intuitions"?)

It may be that only intentional states are subject to the skeptical paradox. After arguing that Wittgenstein denies that there is any inner experience of meaning, Kripke says, "[t]he case is specifically in contrast with feeling a pain, seeing red, and the like."61 Farther on we find:

'Mental states' and 'mental processes' are those introspectible 'inner' contents that I can find in my mind, or that God could find if he looked into my mind. Such phenomena, in as much as they are introspectible, 'qualitative' states of the mind, are not subject to immediate sceptical challenge of the present type.62
This would seem to confirm the view that there is a fact about me (or you) when experiencing +pain+, despite this fact having nothing to do with the meaning (=use) of "pain". Still, unless +pains+ are identical with physical/functional items, these facts cannot be physical/functional facts. Here, if the above is correct, the conclusion to be drawn is that the materialist cannot lean heavily on PLA in hopes of denying that there are purely phenomenal objects of experience.
6. THE ROLE OF EXPERIENCE: PART II

To a considerable extent, so far as many mental states go, by their qualia shall ye know them. To an even greater extent, by their qualia shall ye value them, imagine them, remember them, and fear them.1

On Leaving Something Out

It may seem that science can state the conditions under which people feel pain, or what the state of the core person is when +pain+ is experienced, or what things +pain+ is a response to, but all these, after all, are not definitions of the word "pain", nor conditions essential to our identification of +pain+. These conditions point to facts about the thing, +pain+, and an acquaintance with the referent of "pain" is typically presupposed in these descriptions. What it is about a sensation that leads the accomplished language user to call it "pain", that accounts for the sympathy we feel when we predicate "pain" of another, is the what it is like feature of pain, which is verbally indefinable.

The referent of "pain", a term which figures in our prescientific concept of pain, can only be learned ostensively via experience. To claim, as does Lycan, that "To be in pain of type T . . . is for one's subsubpersonal ð-er to be in characteristic state ST(ð), or for a characteristic
activity $A_T(\emptyset)$ to be going on in one's $\emptyset'$er"\(^2\), indicates the functional/relational characterization of states type-identified as mental states. Construing this to mean that "pain" is defined, or $+$pain$+$ identified solely in terms of causal conditions, shows, I argue, that functionalism fails to adequately capture our concept of pain, or at least that part which we care about.

On Churchland and Churchland's view of it, "Functionalism is to be construed broadly as the thesis that the essence of our psychological states resides in the abstract causal roles they play in a complex economy of internal states mediating environmental inputs and behavioural outputs."\(^3\) As M. T. Thornton remarks, "Our conception of pain may be logically tied to the causal conditions which familiarly obtain, but the concept of pain is not so tied."\(^4\) Felt pain sensations ($+$pains$+$) could (logically) be associated with different causes and effects without foregoing their hurtfulness, without ceasing to be $+$pains$+$. Claiming that purely causal/functional descriptions are ontically neutral (may be true of either mental or physical items) can avoid the charge that our concept of pain is not adequately captured only if the what it is like features of sensations are also ontically neutral. But, I argue, they are not.

Pain is identified as that particular experienced item that mediates between damaging stimulus and pain-behavior,
not as whatever it is which so mediates. Since pain-functional profiles may be realized by different (physical) items in different organisms (or even the same organism at different times), "pain" can be defined as, or pain identified as, whatever it is that plays a certain causal role only if whatever it is that plays a certain causal role is that to which a particular pain is reductively identified. It may be that one neural item in Smith and another in Jones plays the appropriate causal role, but in either case it is +pain+ which must be reduced—prima facie, "pain" refers to +pain+, not merely whatever it is that plays a certain causal role. Our concept of pain is such that "pain" is an ostensive term "defined" (nonverbally) by reference to a particular quality of experience. And any adequate analysis of our concept of pain must reflect this fact. As Thomas Nagel puts it:

Any reductionist program has to be based on an analysis of what is to be reduced. If the analysis leaves something out, the problem will be falsely posed. It is useless to base the defense of materialism on any analysis of mental phenomena that fails to deal explicitly with their subjective character.

If physicalism is to be defended, the phenomenological features must themselves be given a physical account. But when we examine their subjective character it seems that such a result is impossible. The reason is that every subjective phenomenon is
What "point of view"? That which enables us, for one thing, to form an attitude of sympathy (or delight) for another attributed pain. That this "subjective phenomenon", the phenomenon essentially connected with the point of view of an experiencer, is what "pain" is ostensively "defined" by reference to, is what enables us to identify our pains as pains, will now be shown. To stroke one in this direction, it is necessary to show that the role of sensation language, for example, reports of pain, cannot be filled by nonostensive terms having to do with physical/functional characterizations.

Indeed, the role of sensation language can only be analyzed by, so to speak, going beyond language. Physical/functional descriptions or explanations do not amount to complete descriptions of sensations as sensations, for they do not inform us of all that there is to know about, e.g., pain; namely, what it is like to experience +pain+. Nor is the challenge here that the materialist must say in physical/functional language that which cannot be said. The materialist must, however, explain how there can be something which cannot be said in physical/functional language. So, the materialist must give an account in which
purely physical/functional items have subjective character, in which knowing all the physical/functional facts amounts to knowing all there is to know about what it is like to, say, feel pain. If there are features of mental states unknowable to those whose mental states lack such features, functionalism as materialism is untenable. There are such features and science cannot ignore them, for they are integral to understanding human action and experience. Let us unfold this objection.

Suppose we were to have a complete functional description of the neuroanatomical structures upon which a (human) +pain+ in X at t supervenes; assuming this were possible, it would nonetheless not be a complete description (although perhaps as complete as the scientist can give): it would not inform us of all there is to know about the items (or experience) in question. Such functional description would not inform us of what it is like for a person, whose inner goings-on realize such a functional syndrome, to experience a sensation as a sensation. These functional (or, as realized, physical) descriptions can be understood by a person as sensations only if sufficiently similar functional characterizations are, or have been true of (are, or have been realized by) that person. Mere functional description (explanation) is not enough to render a sensation understandable as a sensation, not enough to provide understanding
of, e.g., "pain" ("pain-experience"). Consider more specifics.

Think of the case of Smith, who is congenitally unable to feel pain and the warranted assumption that Smith does not realize the functional profile (future science tells us is) associated with +pain+ (cf. the case of Miss C., pp. 71ff above). Moreover, imagine that Smith has a LaPlacean understanding of the neuroanatomical items that realize the functional characterization of pain. It would seem reasonable, thus, to assume that Smith's inability to feel pain can be (causally) explained by Smith. Smith, however, as long as he has only a functional/physical understanding of pain and does not himself realize pain's associated syndrome of causal happenings (does not experience +pain+), does not understand that to which "pain" ("pain-experience") refers—does not have a complete understanding of (our "concept" of) pain.

Thus Smith is able to provide as complete a functional/physical description and explanation as science will allow of all that happens to Jones, where Jones is like you and me, a feeler of pain. Still there is one fact about what is happening to Jones, which Jones knows (by acquaintance), but which remains completely unintelligible to Smith: what it is like to experience +pain+. Yet, if knowing all the physiological/functional facts about Jones amounts to
knowing all there is to know about Jones, Smith would already have a complete understanding of "pain" as experienced by Jones. But, by plausible hypothesis, Smith does not; indeed, Smith does not have an understanding of that to which "pain" refers. (As has already been hinted, there is, as commonsense and common experience tell us, an object of experience (or at least a quality of that which is experienced) to which "pain" refers. That this "object" is mysterious, does not make it vanish; see below.) To claim that obviously Smith doesn't know this because he hasn't experienced "pain", is to admit the point. Which is? That which is experienced by Jones when he, for instance, touches, ceteris paribus, a red-hot iron at t, is not a physical item. (A conclusion one is understandably loathe to draw; I do not, however, know how to avoid it and still remain satisfied with the explanation of phenomenological features. Nor do I see that denying "pains" thing-status, if plausible, provides a solution; see below.)

A possible response here, which aids in clarifying the problem, is that we can "piece together" what it is like to have sensation S, even though we have never experienced S. Perhaps someone who has experienced S can tell us by comparing it with sensations we have experienced. Although this response does point out a peculiarity with the term "like", it seems to me to amount to a denial of the
ostensive character of "pain"—it should not be surprising that words fail here. Consider Miss C.:

It is interesting to note that while Miss C. was curious about the nature of the feeling accompanying the pronounced pain reactions she observed in others, she apparently had no concept of this herself. (p. ... above)

We cannot describe (other than metaphorically) what it is like to feel pain because of the ostensive character of +pain+; it is not like, in any relevant sense, any other experience. To capture the relevant sense, simply try to imagine how you would describe to Miss C. what it is like to experience +pain+—without leaving something out. What is left out is how it is for the experiencer of +pain+, most of us. We must not be misled into thinking this experience can be described because we can imagine what Morgan's +pain+, at the loss of a limb, was like. Miss C. has never experienced any +pain+ and no comparison (as I can imagine it) will even come close to enabling her to know what it is like.

Now, if one wishes to maintain that they do not understand the locution "know what it is like to experience +pain+", I can do little else but ask them, for example, to touch a red-hot iron; if they can experience +pain+, they will then understand, if not, they will never understand—despite what can be said by way of comparison with other sensations. Any attempt, in the case of +pain+, to piece
together what it is like to experience *pain* will leave something out—that something, the subjective phenomenon, cannot be described.

This subjective phenomenon to which "pain" refers is inaccessible to Smith (and hence, to objective functional/physical theory). Understanding of this feature is attainable, as far as it is attainable, only by an exper- iencer of *pain* (only from this "subjective point of view"). To repeat, no matter how much Smith discovers about the physical aspects of Jones' nervous system (about the core person), Smith will not discover (come to an understanding of) what the experiences Jones had were like (simpliciter, not merely what they were like for Jones). No amount of new information about the purely physical/functional processes will amount to information for Smith about what it is like to experience the sensation of pain. "What it is like" refers to the "object" that is experienced when one experiences a sensation as a sensation viewed from the standpoint of the experiencer. (At this point there is some ambiguity in the notion of pain-experiencer/experience of a sensation. As we shall see, whether we maintain that sensations are (mental) objects, or experiences modes of sensing (a view argued against below), the subjective state lacked by Smith eludes successful materialist treatment.)
A similar case, put forward by Frank Jackson, makes essentially the same point. Consider a brilliant neuroscientist, Mary, who investigates the world in entirely black and white surroundings via a black and white television monitor. Mary's specialty is vision and she acquires all the physical/functional facts that can be obtained about what goes on when we see ripe tomatoes, or the sky, and use terms like "red", "blue", and so on. Jackson asks:

What will happen when Mary is released from her black and white room or is given a colour television monitor? Will she learn anything or not? It seems just obvious that she will learn something about the world and our visual experience of it. But then it is inescapable that her previous knowledge was incomplete. But she had all the physical information. Ergo there is more to have than that, and Physicalism is false.6

Can this conclusion be avoided or rendered harmless by the materialist? Let us investigate.

Attempted Solutions

It will not do to argue, as does Richard Rorty,7 that neuroanatomical predicates (where neuroanatomical items realize the functional profiles of the functionalist) will take over the roles of current psychological predicates. As both Cornman and Bernstein argue:
It is not important, then, which words we use now or ever. What matters is what descriptive roles they play. . . . Thus because "stimulation of C-fibers" taking on the descriptive role of "pain" accomplishes only the elimination of "pain" and not its role in true descriptions, such an elimination of sensation-terms fails to help the eliminative materialist. Indeed, if this is the only way sensation-terms can be eliminated, we should reject eliminative materialism, because we must either keep sensation-terms to make true descriptions or change physicalistic-terms in such a way that using them descriptively implies that there are sensations.

We might concede that neurophysiological discourse is a better way of scientifically explaining and describing the relevant phenomena and I may agree that "sensations" turn out to be brain processes. But if I am to describe these brain processes as I experience them then I must use phenomenal predicates to describe them or if I adopt a new language, the new expressions must at least express what I now express when I report and describe my "sensations."

In reply, Rorty writes: "To back up the claim that it is coherent to suggest that 'the entire descriptive vocabulary of a mode of discourse' might wither away while leaving our descriptive ability undiminished, I can offer no better argument than the example of demons." Since I have dealt with the eliminative claim above, and claimed that there is little plausibility in the claim that +pains+ are like demons, I will rest with the cases of Smith, Jones, Mary, and the above quotes. So, in recognizing the difference between Smith and Jones (between, we might say, the "objective
and subjective viewpoints"), we note that the referent of "pain" is ostensively identified as the what it is like feature of this experience, where that feature exists independently of the words used to describe it. Thus, if a neuroanatomical predicate is to take over the role of, e.g., "My pain is intense", it too must refer to the what it is like feature which can only be picked out, in "an understanding way" (as Jones does), ostensively. But then the neuroanatomical predicate does not take over the role of "pain" so much as absorb, into a new theory-laden way of talking, the fact that its referent is ostensively picked out. Thus the new expressions must account for the understanding Jones has (and Smith lacks) when he says, "My pain is intense"—again, a +pain+ by any other name is a +pain+. Consider how J. L. Mackie views the ontological problem:

It is sometimes argued that language could change so that talk about brain processes and the like would replace talk about experiences. Well, suppose that it could, and did: this would not settle our problem. Part of our problem is whether this reformed language would be adequate, whether it would cover the world, not merely in the sense of reporting and predicting all concrete occurrences under some description, but also in the sense of not leaving undescribed some perfectly real and (to us at least) interesting features. It is argued more plausibly on the other side that whatever discoveries science comes up with, and whatever arguments
physicalist philosophers deploy, we shall always go on speaking in terms of experiences and intentional thought. . . .

For emphasis, I would add: "Speaking in terms of experiences no matter what words we might use." Smith's descriptions (in physical language) of Jones must, if they are to "cover the world" and bring to Smith understanding of "some perfectly real and . . . interesting features", be backed by ostensive identification. Perhaps this problem can be avoided if +pains+, as objects, can simply be shown not to exist.

J. J. C. Smart hints that sensations, as objects of experience, do not exist:

I am not arguing that the after-image is a brain process, but that the experience of having an after-image is a brain process. It is the experience which is reported in the introspective report. . . . So to say that a brain process cannot be yellowy-orange is not to say that a brain process cannot in fact be the experience of having a yellowy-orange after-image. There is, in a sense, no such thing as an after-image or a sense-datum, though there is such a thing as the experience of having an image, and this experience is described indirectly in material object language, not in phenomenal language, for there is no such thing.

Strange; does this entail that there are no +pains+? Smart wants to avoid the claim that the what it is like feature of a sensation, e.g., +pain+, construed as an
instance of a phenomenal property, shows that any (let us say) token-identification of mental items as physical items is doomed because no particular physical item has such a property instance (if it is a purely physical item). If there are no sensations, only experiences (sort of like the brook without the babbling), then there is no thing that has phenomenal properties. (Still, it seems that the experience is an experience of something, and if that something is a neural item, then that neural item does appear to have phenomenal properties. As we shall see below, the introduction of topic-neutral properties at this point does not save materialism from appearances.) Considering the what it is like feature experienced by Jones, but not Smith, as +searing pain+, we see that the following, taken from James Cornman, closely captures our intuitions concerning phenomenal items:

"S" is a phenomenal term =df

(1) The following is true: "There is an (S, instance of S) iff it is experienced by someone, whether or not it is also perceived, and

(2) For any nonlogical term, "P", if "There is an (S, S entity)" entails "There is a (P, P entity)" then

(a) either "P" meets condition (1) or "(Each P, P entity) is a subject of experiences" is true and
Suppose our phenomenal term is "hurting", then by "an instance of hurting" we are to understand a spatio-temporal non-recurrent item which is an aspect of hurting (hurtfulness) but does not exemplify hurting (hurtfulness). For example, my particular pain is said to exemplify the property (construed as a universal) hurtfulness in virtue of this bit of hurtfulness inhering in pain. "This bit of hurtfulness" denotes an instance of the property hurtfulness—the particular hurting of my present pain. Notice also that "hurting" is not a phenomenal term if applied to physical objects since physical objects exist without being experienced or perceived. For someone like Smart, who denies that sense experience consists of relationships between subjects of experiences and phenomenal objects (sensations), one may wish to add, as does Cornman, a disjunct to the first condition: "or each S is an event of a person experiencing in some way but it is not an event of the person perceiving some object." As we shall see, this concession to adverbial materialism leaves us with unexplained appearances.

We see, then, that Smart, in claiming that there are no such things as sensations, seeks to take away the home of phenomenal property instances. Jaegwon Kim pursues this tack:
he [Smart] denies that there are such things as after-images and sense data. And this seems both sound and inevitable. If the basis of the objection from phenomenal properties is correct, the identity theory cannot tolerate mental objects like sense data and pains as legitimate entities in one's world. If they exist and have the phenomenal properties they are assumed to have, the identity theory is false.16

Kim goes on to give an idea of how such "things" are to be avoided:17

A person's having a pain at t is of course not throbbing; and that of course is the whole point of this approach. Rather, throbbing is absorbed into the constitutive property of this event; the event we have is a person's having a throbbing pain (it may be useful to express this event as "a person's having-a-throbbing-pain" to emphasize the fact that on the present approach "throbbing pain" is not a nominal term). It is this event, a person's having a throbbing pain, that is to be identified with a brain event. Similarly, there is no object, a visual image, that has the property of being dim; rather, the event which the identity theory is anxious to identify with a brain event is a person's sensing a dim visual image at a time, where sensing a dim visual image is the mental property the person exemplifies at the time. This property is of course not to be analyzed relationally involving an object, a visual image which is dim, and a person who senses it.18

According to this tack, then, no thing is painful (nothing is painful?), although there are persons who experience painful sensations. Showing that this renders
plausible the claim that sensation, qua (quality-) objects of experience, do not exist is, however, no easy task. Nor does it appear a successful line to follow. According to this approach, which is backed by precious little argument, we are to understand the sentence "There are persons who experience painful sensations" as

1) $(\exists x) (x \text{ is a person and } x \text{ experiences a painful sensation}).$

Intuitively, it would seem that there is something the person experiences, namely, a painful sensation. But the latter would best be captured by

2) $(\exists x) (x \text{ is a person and } (s) (s \text{ is a painful-sensation and } x \text{ experiences } s)).$

Kim's move, however, is to deny that sensations exist; thus he cannot accept 2), for it entails

3) $(\exists s) (s \text{ is a sensation}).$

Typically, the next step is to render experiences as processes or events and 1) as

1) $(\exists x) (\exists e) (x \text{ is a person and } e \text{ is an experiencing-a-painful-sensation and } e \text{ happened to } x),$

where the hyphens preclude the existence of additional logical structure, amounting to the claim that "experiences a painful sensation" (understood as the fused "experiences-
a-painful-sensation") is a predicate with no semantic structure. Surely this is wrong, for if it is true that

4) Jones experiences a searing, throbbing sensation in his back,

then 4) entails

(i) Jones is experiencing a searing sensation,
(ii) Jones is experiencing a searing sensation in his back,
(iii) Jones is experiencing a searing, throbbing sensation,

.. and so on.

To adequately explain such entailments it is best to analyze 4), not along the lines of 1^), but rather

5) (∃x) (∃e) (∃s) (x is a person, e is an experiencing of s and e happened to x and s is a sensation and s is searing and s is throbbing and s is in the back of x.)

Interestingly, however, these intuitive considerations lead us from 5) to its entailment

3) (∃s) (s is a sensation).

Lacking convincing argument to the contrary, the only motivation for analyzing as does Kim, is that sensations are made
to vanish; but so are wanted entailments.

An attempt to escape this reasoning and avoid commitment to sensations as objects, centers on the attempt to reconstrue statements which purport to be about sensations as being instead about the way or manner in which a person is sensing. On this "adverbial theory", "Jones has a pain" becomes "Jones senses painfully." As we have already seen, however, pains have many properties: throbbing, searing, jabbing, etc.; this leads to what Frank Jackson\textsuperscript{21} labels the many property objection.\textsuperscript{22} On the adverbial view, to have a (pain) sensation which is throbbing, is to sense throbbing-ly. With regard to a (pain) sensation which is both throbbing and searing, however, it can be reexpressed as sensing throbbing-ly and searing-ly or sensing throbbing-searing-ly. Taking the former way (a sensation which is T and S analysed as sensing T-ly and S-ly) leads to the difficulty of being unable to distinguish between

6) Jones has a throbbing pain and a searing pain

and

7) Jones has a throbbing, searing pain

(where "has" is understood as "experiences"). For 6) would be reexpressed as
61) Jones senses throbbing-ly and searing-ly, and so would 7). It might be thought that this could be avoided by changing 6) to

8) Jones has a throbbing pain in his foot and a searing pain in his shoulder.

and 61) to


But where "throbbing-in-the-foot-ly" and "searing-in-the-shoulder-ly" are fused, that is, fundamental modes of sensing with no semantic structure, 81) would not and 8) would entail

a) Jones has a throbbing pain
b) Jones has a throbbing pain in his foot
c) Jones has a pain in his foot

... and so on.

Thus, 81) would not be an adequate analysis of 8), for it would not validate the entailments of 8). Nor (if it even makes sense) would

82) Jones senses throbbing-ly and in-the-foot-ly and searingly-ly and in-the-shoulder-ly,
be satisfactory, for it would fail to distinguish between 8) and

9) Jones has a throbbing pain in his shoulder and a searing pain in his foot.

Surely there is no difference between "Jones senses A-ly and B-ly and C-ly and D-ly" and Jones senses A-ly and D-ly and C-ly and B-ly".

For essentially the same reasons, it will not do to analyze a sensation which is T and S, as sensing T-S-ly, for

10) Jones senses throbbing-searing-ly

also fails to distinguish between 6) and 7).

According to the second objection to the adverbial theory, which Jackson labels the complement objection, to have (experience) a sensation which is throbbing cannot be to sense throbbing-ly, for it is possible for both

11) Jones has a throbbing pain

and

12) Jones has a nonthrobbing pain

to be true of Jones at the same time. But then, the conjunction of 11) and 12) would be analyzed as

13) Jones senses throbbing-ly and nonthrobbing-ly (at t).
Given that the adverbial theory fails to account for the intuitive entailments we require for an adequate analysis of sensation statements, we seem justified in claiming that 5), say, entails

3) (s) (s is a sensation).

This, however, is not to deny that 5) also entails

14) (e) (e is an experiencing),

but to reinforce our intuitive conviction that when Jones experiences a sensation, there is, besides the experiencing, something which Jones experiences, namely, a sensation. Thus, if sensations can be shown to have nonphysical properties, then, on the assumption that something with mental (nonphysical) property instances is a mental (not purely physical) object, we are forced to a nonmaterialist view of sensations (if, as I have argued, topic-neutral properties cannot be successfully invoked in the materialist cause).

It looks as if the adverbial theorist wants to identify the sensation of pain with the state of being in pain, in the sense that one has the former if and only if one is in the latter, and since the latter is a functionally characterized state, materialism is home free. The difficulty with this notion turns on the interpretation of "has" (= "experiences"? or "is aware of"?), for the identity, to save materialism,
must be understood such that being in the pain state is constitutive of having the sensation of pain and not merely the cause of felt pain. With this in mind, let us look at another attempt at rendering functionalism as materialism.

Notice that in the preceding objections it is not being claimed that considerations of logical form alone show the adverbial theory wanting. One could, that is, maintain, as does Robert Kraut, that

\[(15)\] Jeremy has a toothache in his upper right bicuspid

is to be given the form

\[(15_1) (\exists x) (\text{toothache, } x) \text{ and Has (Jeremy, } x) \text{ and In (his upper right bicuspid, } x)).\]

This avoids the objections raised above and, it is claimed, commitment to quality-objects as long as we can specify the truth-conditions for \((15_1)\) in such a way that no reference is made to such items—grammar, it might be claimed, is misleading. Kraut pushes this line and maintains that an act-object analysis of experiential discourse need not be embraced, for there is

a semantical account of experiential discourse which draws upon functionalist intuitions about the nature of experience. The account demands no reference to sense-data or other sensory objects, and it explains various inferences permissible in sensory discourse. . . . One of the
most cogent arguments for the existence of sense-data, and of sensory objects generally, is thus undermined.\textsuperscript{24}

Kraut thus accepts (15\textsubscript{1}) as the \textit{correct} logical form for (15); what he must provide is an account in which \textbf{no reference} need be made to quality-objects and which also \textbf{explains} various wanted entailments. The trick is to understand the truth-conditions for (15\textsubscript{1}); thus, given functionalist "intuitions", (15\textsubscript{1}) is true if and only if

\begin{equation}
(15\textsubscript{2}) \text{Jeremy is in that state which would typically be caused by damage to the nerves in his upper right bicuspid.}\textsuperscript{25}
\end{equation}

With regard to the variable in (15\textsubscript{1}), the suggestion amounts to the following:

Initially, let us treat variables which occur within the scope of intensional verbs as ranging, not over entities [actual sense-data] . . . but rather over [functionally characterized] entities . . . which represent the conditions that specify the typical causes of the experiential state in question.\textsuperscript{26}

Or, more to the point:

On the account I propose, what appear to be reference to intensional objects is in fact reference to physical objects which are present in a multiplicity of counter-factual situations--viz., those situations which are relevant to the functional specification of the experiential states in
question. . . . a natural implementation
of a functionalist theory of experience.²⁷

On the assumption that the formal mechanics work out, I'm not at all sure what is accomplished other than the establishment of the claim that if the functionalist account is satisfactory, an adequate semantics is available, i.e., one which avoids the objections lately canvassed—wanted entailments go through. The above discussion of the adverbial theory, although couched in terms of sentences and their entailments, is based on phenomenological considerations (see chapter 2, above) not (as I understand it) addressed by Kraut's formal semantics—unless, that is, one concedes that functionalism provides an adequate account of awareness. I do not argue that functionalism is inadequate because one can't rig truth-conditions for sentences such as (15), but that the truth-conditions that are so generated do not provide sufficient understanding of what goes on when pain is experienced. Being aware of the truth-conditions for being in the pain state, in the sense Kraut intends, ignores what I consider the real issue—for Kraut's phenomenal individuals are taken to be nonexistent intentional objects of mental states (but see chapter 2, above).

When I speak of this +pain+, I am not describing the (physical) object causing my experience, I am making reference to that of which I am aware. The fact that a functionally
characterized neural item is causally related to Jeremy's sensing in a certain way cannot plausibly be described as Jeremy's being aware of that neural item. As I have argued, Jeremy's having a (being aware of his) toothache still has a quality about it that is not explained in terms of typical causes or by providing a functional specification of the experiential state in question—we may be supplied with truth-conditions for being in the pain state, but once in it +pain+ supervenes. So even though it may be true that an experienced state is caused by (functionally specified) neural items, a particular felt pain is not an instance of a physical/functional property. If we ignore the phenomenological data, what is an adverbial theory of? Thus, even if "This pain I feel now is intense" is true (as uttered by me) if and only if I realize a certain functional characterization, it does not follow that an act-quality-object (or act-particular, or act-element-of-being) analysis of experience is not required. Again, at most Kraut establishes that the functionalist program can be given a defensible semantics only if the truth-conditions for experiential discourse are adequate to the phenomenological data. It does not appear that they are (though, I admit, conclusive proof here is not available). Still, for reasons given above (especially chapter 2), I deny that Kraut establishes his claim that "what appears to be
reference to intensional objects is in fact reference to physical objects". By changing the semantics, the functionalist can claim that certain marks entail other (wanted) marks; but it is not (mere) marks that are wanted, it is what they refer to that "creates" the wanted entailments—the cost of "grammaticalness" (standard logical form) is that the theory no longer fits the phenomenological data adequately or plausibly.

Finally, as I understand Kraut, he is faced with a dilemma. Consider his proposal:

Let us treat "is an after-image" as an intensional predicate, by the following rough criterion: any world in which a sentence of the form 't is an after-image' is true is a world which contains a sentient being in a specific state of awareness. Bound variables . . . occurring within intensional contexts, thus range not over entities present in the actual world (e.g., after-images) but rather over world lines, the values of which are entities present in those worlds which represent the conditions that specify the typical causes of the experiential state in question.

Thus reference is not to actual yellowish-orange mental objects, but to merely possible yellowish-orange objects (a class of world lines). The first horn of the dilemma has to do with the reification of possible worlds. In short, his apparatus of worlds, neighboring worlds, world lines, etc., is, at best, mysterious. Indeed, Kraut seeks to avoid this
horn by not reifying possible worlds and he maintains that "[p]ossible worlds discourse is, in what follows, a mere façon de parler for counterfactual discourse."^®

This move sets Kraut securely on the second horn of the dilemma, for now he must maintain that, for example, the sentence "I see a yellowish-orange after-image", is to be analyzed counterfactually as sensing as you would be if you were [public event]. The bottom line of taking this horn of the dilemma is simply to retreat to a Smartian topic-neutral analysis, an analysis I have already shown flawed.

In keeping with these considerations, let us look quickly at an argument David Wiggins levels against Kripke's (I will assume familiar) attack on the Identity Theory.

Wiggins begins:

One argument which I piece together from Kripke himself for (6) is this:

(8) Heat is the cause of the sensation of heat

(9) Molecular agitation is the cause of the sensation of heat

". (6) [Heat is molecular agitation].^®

Wiggins goes on the attack:

The trouble is that there seems to be a parallel argument for materialism

(10) pain is what affects us painfully

(11) neural state n is what affects us painfully
\[I\]Being in pain at time \(t\) is being in neural state \(n\) at time \(t\).\]

(Wiggins here assumes Kripke’s: \((x) \ (6) \ ((x = y) \to \square (y = x)))\). The obvious problem here is with (10) and the claim that "pain is what affects us", for this appears similar to Kraut’s causal analysis. Yet, as I have argued, in order to capture the phenomenological data, (10) should be understood as

\[(10)\text{ pain (+pain+)} \text{ is (identical with) our being affected painfully.}\]

But then the proper conclusion is not (7), but (given my causal understanding of "what affects us")

\[(7_1) \text{ Being in pain at time } t \text{ is caused by neural state } n \text{ at } t.\]

Consider the functionalist claim:

\[(A) \text{ My being in pain at } t = \text{ my being in psychological state } n \text{ (of the } S_p \text{ kind) at } t.\]

To avoid Kripke’s argument against such an identity, the functionalist must (it seems) claim that my pain at \(t\) is identical with whatever state of me realizes functional syndrome \(S_p\) at \(t\)—thus "pain", so used, is not rigid. Not wanting to deny the obvious, let us assume that pains are linked with feelings of certain sorts. We may then go on to claim
(B) My physiological state \( n \) \((\text{of the } S_p \text{ kind})\) at \( t \) feels painful.

Suppose we now adopt an adverbial notion of pain and claim

(C) My being in pain at \( t \) is my being painfully affected at \( t \).

It would seem to follow either that (using Wiggins' terminology)

(D) Physiological state \( n \) \((\text{of the } S_p \text{ kind})\) at \( t \) is what affects me painfully at \( t \)

or

(E) Physiological state \( n \) \((\text{of the } S_p \text{ kind})\) at \( t \) is my being affected painfully at \( t \)

will be needed to generate (A). But, as we have just seen, it will not do to claim that being affected painfully is what pain causes and then go on to identify my particular pain at \( t \) with my being in physiological state \( n \) \((\text{of the } S_p \text{ kind})\) at \( t \). This will not capture what the materialist needs to capture, viz., my feeling pain or being painfully affected; all this amounts to, given (D), is the claim that physiological state \( n \) \((\text{of the } S_p \text{ kind})\) causes my feeling pain. It seems, then, that we should opt for (E). But this does not aid the materialist, for physiological state \( n \) \((\text{of the } S_p \text{ kind})\) has not inherited the same features that are essentially nonphysical aspects of pain.
So, given the non-rigidity of "pain", it would seem to follow that

(F) My pain at t might have failed to be a pain at t,

that is,

(G) My physiological state n (of the S_p kind) at t might have failed to be a pain at t.

However, the functionalist would want to deny (G), for "my pain" in (F) (given (E)) does not simply refer to physiological state n (of the S_p kind), but rather to physiological-state-n-(of-the-S_p-kind-) feeling-painful: it is my being painfully affected that is to be accounted for. But then, (G) would read

(G_1) My-physiological-state-n-(of-the-S_p-kind-) feeling-painful at t might have failed to be painful.

(G_1) is absurd. But to avoid this, "my pain" must be construed as rigid, and thus, (F) and (G) false. If so, the functionalist owes us an account of the necessity in:

□¬(G), i.e.,

(H) □ (my physiological state n (of the S_p kind) at t is a pain at t)

But, as we have seen, such an account is difficult to come by, since (A) is synthetic and a posteriori. It is much
easier (for me) to understand, necessarily: my being in pain at $t$ is caused by my being in physiological state $n$ (of the $S_p$ kind) at $t$. Yet this is perfectly compatible with a nonmaterialist view. Has the functionalist "slid" from the plausibility of this latter claim to the plausibility of (H)?

**Quasi-Epiphenomenalism**

An interesting attempt to incorporate this "dualistic" theory of sensations within a materialist (or, perhaps, "quasi-materialist") framework has been put forth by George Wilson, who espouses a quasi-epiphenomenalism. His view is the conjunction of four theses:

There are experiences of sensations . . . and they are identical with neural processes. . . . There are also sensations . . . and they are nonphysical entities. . . . What I have been calling "quasi-epiphenomenalism" has been designed to demonstrate that this conjunction is perfectly consistent and suffers from no obvious conceptual incoherence.

We can understand Wilson's position by considering our earlier diagram of the epiphenomenalist's position:

```
  +P+
   ↑
  I---N1---N2---N3---A
```
Where the solid line represents the causal path leading from injury, I (of an organism), to avoidance behavior, \( A \), \( N_1 \) will then represent the neural item that is the proximate cause of the pain sensation, \(+P_+\), a purely mental item. Wilson's move is to claim that experiences of sensations are identical with those neural items (say, \( N_1 \)) that directly cause sensations \(+P_+\). Thus,

\[
\text{distinguishing between sensations and experiences of sensations, it is possible to hold that although the experienced pain \ldots is never a cause of behavior, the experiencing of that pain \ldots may be. \ldots what thereby happens to someone experiencing a sensation is the experiencing of that sensation, and we can go on to propose that it is this happening, this event, which, strictly speaking, has the causal properties we ordinarily ascribe to sensations. In this way, quasi-epiphenomenalism seems not so much to conflict with as to refine upon our commonsense beliefs about sensations and the causation of behavior.}^{34}
\]

For Wilson, then, "all behavior can be explained in purely physical terms" (emphasis mine).\(^{35}\) What strikes me as problematic, however, is that these sensations would not be explained in purely physical terms and would "dangle"—all that happens to an experiencer would not be accounted for. Thus, although experiences of sensations, construed as identical with neural items, would not be, sensations would be nomological danglers; as Feigl warns:
It [epiphenomenalism] accepts two-fundamentally different sorts of laws—the causal laws and laws of psychological correspondence. The physical (causal) laws connect the events in the physical world in the manner of a complex network, while the correspondence laws involve relations of physical events with purely mental "danglers." These correspondence laws are peculiar in that they may be said to postulate "effects" (mental states as dependent variables) which by themselves do not function, or at least do not seem to be needed, as "causes" (independent variables) for any observable behavior.36

Although such an objection would not seem conclusive with regard to epiphenomenalism (proper),37 it does seem to rule out materialism understood (à la Cornman) as:

\[ M = \text{Every existing item is one that has a property only if: (a) the property is either physical, or topic-neutral; or (b) each actual instance of the property is nothing but an instance of a physical or topic-neutral property.}38 \]

As Wilson concedes, sensations, and hence their properties, are nonphysical. Moreover, since phenomenal properties are properties only of sensations, they cannot be topic-neutral (had by items either material or mental). Hence, if sensations are truly anomalous, their being so would take much of the spirit out of materialism (= M)—to put it mildly. Though Wilson agrees that "[n]o doubt the realm of the mental is unique in various ways, . . . it is suspiciously ad hoc and inexplicable that it should be uniquely in this realm
that we should find entities lacking all causal efficacy,"\textsuperscript{39} he goes on to deny that nomological danglers are unique to the mental realm but are present in the physical realm as well. He claims, then, that nomological danglers are not particularly deleterious to his quasi-epiphenomenalism (within a materialist framework), for they are part and parcel of many legitimate physical explanations (often a strategy of the materialist).\textsuperscript{40} Wilson’s (ill chosen) example is that of a rainbow:\textsuperscript{41}

Physicists describe the underlying physical processes that are said to "produce," "give rise to," or "cause" rainbows. . . . Rainbows, we are told, are produced by the diffraction of light rays passing through collocactions of water molecules suspended in the atmosphere. In fact, since no intervening causal process is postulated, we can suppose that the diffraction directly causes the occurrence of the rainbow. The rainbow itself, on the other hand, is not treated as a cause or causal factor of further events.\textsuperscript{42}

Wilson maintains that since rainbows cannot be strictly identified with anything in the physicist’s austere vocabulary, "we are concerned here with a close analogue to the phenomenal properties objection."\textsuperscript{43} Indeed, as I view it, closer than he realizes, for we do not have a nonmental case analogous to a mental case: rainbows (seen colors) are not "public" nonmental entities any more than +pains+ are. Wilson may be thinking that rainbows are public in the sense
that different observers can see the same rainbow. But this is just wrong; we can be said to experience the "same" rainbow in just the sense we can be said to experience the "same" pain. Compare Mackie:

... where we see the lightning there is nothing but an electrical discharge and photons being sent out from it; but we see it as a bright flash. Again the brightness is shifted from the external object to become part of the content of consciousness. And the same with heat as we feel it. But it is obvious that this method of explaining things away will not do for experiences themselves and their contents. Its looking thus and so to us is one of the features that is embarrassing for the physicalist, and the explanation that it only looks to us as if it looks thus and so is not only absurd but useless: it would leave us with an unexplained item of the same sort on our hands.

The phenomenal and intentional features cannot plausibly be denied, and once they are recognized even as merely apparent they cannot be explained away. The only available method of explaining apparent features away fails here through circularity.44

Just as different observers can expose themselves to the same collocation of water molecules, under specified conditions, and experience (ceteris paribus) a rainbow, different observers can expose themselves to (touch) a red-hot iron and feel (ceteris paribus) the "same" pain: neither rainbows nor +pains+ exist unless there are sentient creatures who experience them. In a world in which there are no sentient creatures, there may well be diffracted light rays, suspended
water particles, and red-hot irons (molecular activity of such and such a sort); there will not, however, be rainbows or +pains+ (cf. the definition on pp. 238-9 above).

Thus, Wilson fails to establish his conclusion that "the charge that epiphenomenalism ascribes to sensations a unique and unprecedentented (sic) metaphysical status, is, at best, a charge 'not proven'." Wilson fails to show sensations are not anomalous and M not damaged. Even granting Wilson the contentious claim that experiences of sensations = neural items (construed as the physical tokens that realize the relevant low-level functional types of the appropriate functional profile) and hence, that it is possible to provide a complete causal explanation of the behavior of sentient organisms in purely physical terms, will not enable him to claim that all that happens to sentient organisms (who, e.g., touch a red-hot iron) can be so explained: experiences have content of which sentient creatures are aware and which is not completely explained (covered) in purely physical terms. Again, what it is like to experience a rainbow or a +pain+ is left out by any such complete causal account of behavior.

By way of caveat, let us briefly consider functionalism.

The above serves to set the stage for another functionalist reply:
You ask what pains are, I respond: to be in pain is to realize a certain functional profile. Realizing that functional profile is constitutive of being in pain. True, Smith does not experience pain (Jones does), but he knows what pain is in virtue of knowing all the physical/functional facts. Consider the accepted identity: lightning is nothing but electrical discharge. The brilliant scientist can know what lightning is, know all the electrical facts, without ever seeing a flash of lightning. That he has never seen a flash of lightning, does not "know what it is like", does not damage the identity that lightning just is electrical discharge. Why think the Smith-Jones case of pain any different?

Here, the functionalist is on the attack; but the challenge can be met. Mackie has already pointed out the difference. It is, in effect, that the appearance properties of lightning—the (seen) flash in the sky—have been "transferred" to sensations and left to the philosophers to reduce. Particular instances of +pain+, like particular instances of lightning flashes (+lightning-flashes+), remain unreduced. The flash of lightning, the appearance, is known only to one who has experienced it, despite total knowledge of electrical facts. Lightning may be identical with electrical discharge (interpreted as the fact that there is lightning if and only if electrical discharge), but it is not the case that its being nothing but electrical discharge shows +pain+ is nothing but a realized functional syndrome, for the instances of the phenomenal property remain unreduced.
(they are transferred). Hence, pain may be "identical" with realized functional syndromes, but +pain+ remains unreduced. The analogy appealed to is faulty (I expand on this below).

One may swipe at this claim by maintaining that it entails that if there were no people, there would be no lightning. But this, of course, depends on one's concerns. I maintain only that (trivially) if there were no people (functional systems of such-and-such complexity), there would be no seen flashes, no appearances, no phenomenal property instances. There would still be electrical discharge of such-and-such a sort. If one thinks of the latter as lightning (a most plausible construal), then there would be such without people; if the former (an implausible construal), then without people there would be no lightning. But I agree that lightning is electrical discharge in just the way I agree that being in pain is being in physiological state of the $S_p$ kind—still all there is to being in pain is not thus explained. When aware of our state as a pain state, we are aware of a phenomenal property instance that itself cannot be reductively identified with a realized functional syndrome. Lightning is nothing but electrical discharge, but the seen flash is part of the content of consciousness (cf. Mackie, p. 260, above). Being in pain is realizing a functional syndrome, but the felt pain (that of
which we are aware) is not, although it is caused by the realization. The seen flash is also caused by electrical discharge and photons being sent out from it, but the seen flash is not the latter. Thus, Smith's not knowing what it is like to experience (be aware de re of) pain does not damage the identity: being in pain = being in physiological state \( n \) of the \( S_p \) kind. But this only gives Smith knowledge of what causes \( +\text{pain}+ \), it does not tell him what \( +\text{pain}+ \) is or enable him to understand what "pain" refers to, etc.

The way in which we each identify our pain experiences and the way in which the scientist establishes correlations between felt pains and realized functional syndromes adds fuel to this issue. Whenever we identify our own experience of pain as a pain experience, we do so (once we are accomplished users of sensation language) by noting the what it is like feature of that experience, viz., \( +\text{pain}+ \)--we need no machine readings. Again, no such feature to the experience, no experience of pain. Whenever a neuroscientist (of the future) identifies a particular neural item as a particular experience of pain (makes the claim "neural item \( \mathcal{V} \) (of the \( S_p \) kind) in \( x \) at \( t \) = experience of pain in \( x \) at \( t \)"), the scientist must also indirectly (but ineliminably) support his identification by the what it is like feature, for there is nothing in the neural items themselves (stripped of past correlations with reports of experiencers of pain) that
serves to back the identification of \( \psi \) as a pain experience. Only the correlation of physical items with reports of experiencers of pain (I assume mere (non-verbal) behavior will not suffice\(^47\)) can sustain the identity claim; yet such reports depend essentially upon the experiencers' awareness of the what it is like feature of the pain experience.

In this sense, "pain" is essential to the identification of an experience as a pain experience; likewise the identification of neural items as an experience of pain (not merely the cause of that experience) depends essentially (albeit indirectly) upon this same feature. That is, \( \emptyset \) (experience of pain at \( t \)) can be identified as a pain experience only via \( \alpha \) (felt phenomenal features), while \( \psi \) (neural item) can be identified only via \( \beta \) (physical features). Still, in order for \( \emptyset = \psi \) to be true (where "=" is interpreted as reductively identified), \( \emptyset \) and \( \psi \) must be capable of being identified as the same thing, viz., \( \psi \); but this cannot be done for \( \emptyset \) is always (ostensively) identified via \( \alpha \) and never \( \beta \) (and \( \alpha \neq \beta \)).

It might be thought that the above identity is analogous to the identity "water = \( H_2O \)", which is true even though water is typically identified by the description "the liquid which is wet, cool, etc." (or some such) and \( H_2O \) by its physiochemical properties—thus, by analogy \( \emptyset = \psi \) can also be true. But the cases are not analogous for the
essential property of water is $\text{H}_2\text{O}$, but the essential property of pain is $+\text{pain}+$. So, where $\alpha$ is essential to the identification of $\emptyset$ as a pain experience, $\emptyset$ is a nonphysical item (has $\alpha$ as essential properties), and since to identify $\Psi$ as an experience of pain requires essential use of $\alpha$ (-reports), it follows that $\Psi$ would be nonphysical, contrary to hypothesis. Whereas we might say "$=$" in "water $= \text{H}_2\text{O}$ sides with $\text{H}_2\text{O}$, its properties being essential to water," $=$ in "$\emptyset = \Psi$" sides with $\emptyset$, its properties ($\alpha$) being essential to pain experience; so, if $\emptyset = \Psi$" is true, it is because $\Psi$ has nonphysical (phenomenal) features, again contrary to the hypothesis of reductive identity.

An apparent way out, for Wilson, is to claim that $\Psi$ is identified as $\emptyset$ because $\Psi$ causes the sensation of pain; thus as long as $\Psi$ causes the sensation of pain it can be identified with the experience of pain. I don't see that this changes anything, for still could not be identified as the cause of pain except via $\alpha$ -reports. Even if this could be avoided, the claim would then be that for any $\Psi$ and $\emptyset$, such that "$\Psi = \emptyset$" is true, $\Psi$ must cause the sensation of pain, for if $\Psi$ does not cause the sensation of pain, there is no way to identify $\Psi$ as $\emptyset$ except as above. But this would entail that, where "$\emptyset = \_\_\_$" is true, there could not be a $\emptyset$ without a $\Psi$ which causes the sensation of pain;
in other words, there could be no experiences of pain by, say, a disembodied "mind", since there would be no to cause the identifying sensation. Surely this is an empirical question not to be fixed by the nature of one's analysis. This points to (somewhat of) an advantage for the functionalist, for if the functional characterization of the experience of pain is realized by nonphysical items, functionalism is not thereby shown false, only materialism.

Quasi-epiphenomenalism, then, leaves idle, unexplained sensations dangling, thereby showing materialism false—or that, plus it relegates felt phenomenological features to the status of unnecessary evolutionary accidents. In any event, quasi-epiphenomenalism does not provide a complete account of sensations as sensations (or experiences as sensation experiences)—left out is what it is like to experience a sensation of, say, pain. Finally, let us turn to further phenomenal property objections, and consider more carefully functionalist responses.

Appearances—A Hole in the Functionalist Net

As we have already seen, the problem of phenomenal properties stems from our intuitions that even if mental items (e.g., pain sensations) are to be picked out by their causal properties, or functional roles, where the items having such causal properties or realizing the appropriate
functional roles are physical items (and with regard to the particulars, tokens, not types, they must be physical items), the what it is like feature of, say, pain is left out, unexplained. That is, even if pain is what has certain typical causal relations, the referent of "pain" can only be understood ostensively by experiencing +pain+. Thus, what it is like to experience pain is the mental aspect requiring explanation. Going back to Smart, we find him putting Black's objection thus:

\[\ldots\text{it may be possible to get out of asserting the existence of irreducibly psychic processes, but not out of asserting the existence of irreducibly psychic properties. For suppose we identify the Morning Star with the Evening Star. Then there must be some properties which logically imply that of being the Morning Star, and quite distinct properties which entail that of being the Evening Star. Again, there must be some properties (for example, that of being a yellow flash) which are logically distinct from those in the physicalist story.}^{48}\]

Ned Block comments on Smart's problem:

Smart worried about the following objection to mind-body identity: So what if pain is a physical state? It can still have a variety of phenomenal properties, such as sharpness, and these phenomenal properties may be irreducibly mental. Then Smart and other identity theorists would be stuck with a "double aspect" theory: pain is a physical state, but it has both physical and irreducibly mental properties.\(^{49}\)
As has been alleged, these features are needed to explain (it seems) the causation of pain-behavior, the identification of sensations as sensations and our understanding of the referent of "pain"—in short, these features, understood as phenomenal features, play a role that mere physical features cannot fill. We are now granting that to be in pain is to be construed as sensing painfully. Block seems to view this objection as being aimed at an identity theory, whereas it is, I believe, more comprehensive, being aimed at materialism generally. One version of the above objection can be construed as follows:

1) Sensations are identified as neural items.

2) If (1) is true, then: (a) there is at least one property \( P_1 \) such that something's having \( P_1 \) entails that thing's being a sensation, and (b) there is at least one property \( P_2 \) such that something's having \( P_2 \) entails that thing's being a neural item but neither entails nor is entailed by that thing's having \( P_1 \).

3) If (2a) and (2b) are true, then \( P_1 \) is not a physical property.

THEREFORE

4) Sensations have at least one nonphysical property, namely, \( P_1 \)

5) If sensations have at least one nonphysical property, then materialism is false.

THEREFORE

6) Materialism (\( M \)) is false.
A caveat is in order lest one make the unwarranted assumption that a (materialistic) functionalist need pay little attention to this argument since premise one involves the mind-brain identity thesis. As mentioned, the above argument is properly understood as addressing materialism, and a functional thesis which maintains a token-identity thesis in which functional types are realized by physiological items reducible, in the sense of strict identity, all the way to the subatomic level, falls within its scope. Thus, even if what plays the role of +pain+ in one organism is physical item $N_1$, while in another organism it is $N_2$ which occupies this role, it would still follow that each particular occurrence of +pain+ is some $N_0$ or other. With this understanding, then, the present argument, suitably enriched, aims at the functionalist qua materialist.

The above version of the phenomenal property objection, however, will obviously not do. Recalling that topic-neutral properties (which are not physical properties) may be had by items either physical or mental, premise (5) is seen to be false, for sensations surely have such topic-neutral properties. Their having topic-neutral properties, however, would not show materialism false. Another problem is that, as the argument is rendered, concern seems to be with properties construed as abstract objects (universals) and not with particulars--instances of properties, for
example, the particular hurting of my present pain. Another version of this argument can be gleaned from Smart's further construal of Black's objection:

... it might be thought that the objection succeeds at one jump. For consider the property of 'being a yellow flash'. It might seem that this property lies inevitably outside the physicalist framework within which I am trying to work (either by 'yellow' being an objective emergent property of physical objects, or else by being a power to produce yellow sense-data, where 'yellow' in this second instantiation of the word, refers to a purely phenomenal or introspectible quality. 52

Transferring discussion from "yellow" to talk of "pain" and centering concern on particular instances, we see that this version of the objection is getting close to what has been maintained above. Let me give the skeleton of this argument, attempting to capture the notion that we are concerned with particular instances, particular yellowings or hurtings. Should the skeleton prove inadequate, it is hoped that the following discussion will flesh it out. Consider, then:

1) A particular sensation S in Jones at t instances phenomenal property P (e.g., hurtfulness).

2) No instance of a phenomenal property is an instance of a physical or topic-neutral property.

3) S is identified as neural item N.
4) There are neural items, in particular, N.

THEREFORE

5) There are sensations, in particular, S.

6) There are instances of phenomenal properties, in particular, P.

7) If (6) is true, materialism is false.

THEREFORE

8) Materialism is false.

If this talk of instances makes sense, and I maintain it does, the entire argument rests on premise (2). If the particular +pain+ of which I am aware now cannot be reductively identified as an instance of a physical or topic-neutral property (cannot be explained away), materialism is muddled by appearances. I have argued this cannot be plausibly accomplished via the topic-neutral tack, the PLA or the adverbial theory.

How might the functionalist reply? Suppose, in science fiction style, we are able to look inside Jones and discriminate the various levels of organization, etc., etc.; how will we identify a (say) neural item as realizing a pain-functional syndrome, as opposed to a tickle-functional syndrome? For the functionalist, individuation of mental states (types) is according functional role. So we must have a general account of what makes it appropriate to render a particular functional state (an abstract "program") as
being (when realized) a feeling of pain, etc. It is here, it seems, that the functionalist must turn to behavior, especially verbal reports of experiencers, to establish relevant correlations. Which of Jones' physiological state-tokens are the mental ones of the various mental kinds is determined by functional (behavior-causal) role. Even the notion that the function of one mental state may include the production of further mental states cannot avoid the tie to behavior, for the significance of the former mental state depends on the behavioral tie of the latter. Verbal reports are essential bits of behavior for establishing correlations. Suppose we render the functionalist claim thus:

\[
A \quad \text{Jones' pain at } t = \text{whatever physiological state Jones is in at } t \text{ that is of the } S_p \text{ kind.}
\]

Since "\(S_p\)" simply denotes the relevant functional state realized by physiological tokens, it must be determined that it is of the relevant kind. Since we are not mere behaviorists, we must come up with an account of the particular instance of hurtfulness of which Jones is aware when he identifies and discriminates a pain from a tickle experience. Only first-person statements about mental states reflect reliance on our introspective awareness of how they feel. It might be thought that we can interpret our reports of subjective experience as topic-neutral and claim that all we
know when we report our experiences is that something, we
know not what, is going on which is like what is going on
when [public event]. \(^{53}\) We have seen that this tack is not
successful. Suppose that what we experience are neural
items—do we not know what it is like to experience a neural
item (identified as a pain)? We will return to this ques­
tion later.

On the view we are considering (A), we generate type-
identity between functional syndromes and pain only if we
first establish correlations—if we are to identify exper­
ience of pain as experiences of neural items (or sensations
as neural items), we must begin with correlations. After
all, why think functional syndrome F is a pain-functional
syndrome if it is not correlated with what we experience
(when we, e.g., touch red-hot irons)? Of course this way of
phrasing the question may seem biased, for the materialist
is claiming, it seems, that what we experience when in pain
is a neural item. But, since that appears obviously false
(unless neural items have phenomenal properties and the
appearances of such can be reduced), the materialist cannot
simply announce it (and, in effect, put the cart before the
horse). Correlations must be established between the
realization of F and the reports of pain experiencers.
Correlations with such reports are needed, as alluded to
above, because, when we think in terms of first-person, not
merely third-person concerns, we realize that such reports have accompaniments of which we are aware.

Correlations, however, can be established only if there is individuation/identification of, for example, pains from tickles by experiencers. Experiencers of pain, of course, identify pain via "appearances", not via functional roles or neurophysiological items. So, it would seem that the materialist must maintain that the particular appearance instances must be reductively identified with instances of physical or topic-neutral properties.

One possibility is to, so to speak, interpret the correlations between pain-reports and realized functional states as that of identity. That is, when Jones reports that he has a searing pain, he is actually referring to a neural item as it appears to him. Thus, the thing identified by Jones via his experience is identified with what the neuroscientist detects—we have identifications of one and the same thing. If this is plausible, we have avoided (one sort of) mental entity—but we still have appearances to worry about. There is, then, an unreduced notion of (the objects of) awareness with which we still must deal.

It would seem that the materialist must claim that these appearances are, in some sense, merely illusory. He must deny that we have pre-scientific knowledge of what mental states are (or are like). All we know is that there
is something going on in us which is like . . ., and thus our first-person reports are topic-neutral, mere reference-fixing heuristics—it is left to science to determine what mental states actually are. This is not plausible when dealing with appearances.

If we think of these topic-neutral reports as temporary expedients, mere stopgaps, while we await the deliverances of science, then we will hold, given a faith materialism, that science will show that what these reports make reference to will be physical items. However, as we have seen above, this topic-neutral reference-fixing notion is not plausible, for there is reason to hold that reference to phenomenal features cannot be eliminated—and reductive identification ("nothing but") requires this (at least in principle). Let us imagine the future through the materialist's "eyes" and consider reports of (future) experiences.

Imagine that Paul touches a red-hot iron and avows:

"Neural state \( N_{44} \), which is of the \( S_p \) kind, was just activated." Pat hears this and applies salve to his hand. We can imagine the "conceptual framework" of this future generation changed so that:

They do not feel common objects grow cooler with the onset of darkness, nor observe the dew forming on every surface. They feel the molecular KE of common aggregates dwindle with the now uncompensated radiation of their energy starwards, and they observe the accretion of reassociated atmospheric \( H_2O \) molecules as their KE is lost to the now
more quiescent aggregates with which they collide... They do not warm themselves next the fire and gaze at the flickering flames. They absorb some EM energy in the $10^{-5}$ m range emitted by the highly exothermic oxidation reaction, and observe the turbulences in the thermally incandescent river of molecules forced upwards by the denser atmosphere surrounding.

And, we can imagine, their first-person reports are given in such language. But how does Paul know (what justifies him in his claim that) he was (or is) in such a state? Since we are imagining future science, it would seem that scientific theory justifies the claim, as it does in the above quote—it is how Paul learned the language. But when Paul accidently touches the red-hot iron, what is he immediately aware of, what does he immediately refer to by his avowal? What "triggers" his avowal, informs him that he is in such a state?

All the scientists of the day may respond, given total knowledge of Paul's neuroanatomy: "Neural state $N_{44}$, which is of the $S_p$ kind" (and "$S_p$ kind" can be spelled out in as much detail as we may wish). Still, if that neural state did not give rise to the (what we call) experience of pain (+pain+), if it did not feel like something, Paul would not have known he was in that state (we can even assume that he may not have been)---a +pain+ by any other name... I suppose the scientists of the day can say that the +pain+
Paul experiences simply plays a dispensable role in that, while it aids in the confirmation of certain knowledge claims, it is not indispensable for the description or explanation of reality. But why think this? It seems but dogma to maintain that only language couched in physical terms reveals to us the nature of reality. It is implausible in the extreme to think appearances (e.g., +pain+) can be eliminated and man successfully adapt to the world—implausible in the extreme given the nature of the human organism as we know it. If we were different creatures, all bets would be off. As it is, however, it is in the nature of things that pains hurt.

Invoking topic-neutral statements as stopgaps points out that materialists believe that "mental" statements really refer to physical (neural) items and that "phenomenal" language is but an imperfect way, due to our ignorance, of describing a reality which is better (?) accounted for by the (future) scientific framework. But, though Paul may change his language to physicalistic language (better: have it changed for him), he still (as an experiencer) makes identifying reference via "appearances". I do not see how such appearances can be eliminated or explained (away). What "Neural state N_{44}' which is of the S_{p} kind" refers to is the cause of (or that which is correlated with) +pain+, it does not refer to that which is identified as "pain".
Although implausible, it may even be maintained that "a person's bodily parts ache when their nerve endings are affected in certain ways, regardless of whether the person has an experience of achiness". Left unexplained, however, is achiness I feel when I do experience a, say, toothache—the what it is like feature dangles.

It seems, then, that if we overlook (for the sake of argument) the earlier claims of this chapter and maintain that there are no sensations which we experience—we experience neural items—it may follow that science will tell us what must be true for us to be in the pain state, namely, realization of neural items of the $S_p$ kind. But, as we have seen, it does not follow that we do not know something essential about the pain state when aware of $\text{+pain+}$. While there are many instances of topic-neutral statements, some even involved in reductive identity claims, like "gene = DNA (or RNA) molecule", they are not successful in the case of felt pain—we do know what pains are like (construed as sensations or neural items). As Mackie points out:

The gene is in a different position. It is initially introduced only as a hypothetical entity, as that which plays a certain causal role, interacting in regular ways with other genes, in determining the inheritance of characteristics. To start with, it is simply that, whatever it may be that does such and such. It is therefore freely up for
identification with whatever may later be discovered to do this job. There is no appearance here to be explained away. If mental features were analogous to genes, if they were introduced simply as otherwise unknown causes of behavior, as the grounds, whatever they may be, of dispositions, there would be no obstacle to reductive identification.56

As I have set this difficulty, a particular pain (+pain+) cannot be given a topic-neutral analysis, for a particular pain is identified as a pain in virtue of a particular instance of a felt quality or what it is like feature we experience. Thus, while "gene" may refer to whatever it is that causes hereditary characteristics, "pain" does not refer to whatever it is that causes or accompanies pain-behavior, else Smith would know all Jones knows. "Pain", we may say, refers rigidly to +pain+, be that a mental object or an instance of a phenomenal property (denying that the latter is a mental object seems mere verbiage).

We have seen, then, that we can (and do) identify and know what particular pains are like without knowledge of the physicalist/functional properties alleged to constitute them. Given the ostensive character of pain, it is true that we cannot describe a pain's hurtfulness to one who has never experienced pain; yet we can identify and discriminate pains and know what hurtfulness is without being able to say what it is (for "pain" to be understood is for +pain+ to be felt.)
Finally, consider another case that points out that, even if pains are neural items of the $S_p$ kind, we are left with an unreduced notion of awareness due to appearance properties. Suppose a neuroscientist of the future is able to selectively stimulate my brain so that it (and I) react as if I had touched a red-hot iron. I experience a +searing pain+. Certainly the brain is not searing, but this does little for appearances for it is just wrong to say it is as if I were aware of +pain+, but I am not (really) aware of +pain+. I am feeling pain, caused by the neuroscientist. But now switch to caused color experiences. Here the neuroscientist stimulates my brain and what is produced is an experience such that it is as if I were aware of an instance of yellow. Our materialist must claim that I am not aware of an instance of yellow, it is mere illusion since there is nothing yellow in the room, including my brain—"yellow" refers to a nonexistent intentional object, or my being aware of yellow is my being aware of a neural item. This is a bizarre response: I am not aware of yellow, although it is as if I were aware of yellow? But just as I do feel pain when so stimulated (it hurts!), I am aware of yellow (and a rather pleasant shade at that!). Once again, these appearances cannot be ignored without leaving something out.
Thus, A may be a formula for how to create a certain experience—just organize the physiological tokens in such and such a way—but it does not tell us what particular instances of hurtings or yellowings are. Nor is it plausible to reductively identify them as neural items via stopgap topic-neutral descriptions, for our awareness of the hurt or the yellow provides us with (non-topic-neutral) knowledge. We return to Mackie's point that "the explanation that it only looks to us as if it looks thus and so is not only absurd but useless: it would leave us with an unexplained item of the same sort on our hands."\(^{57}\)

Let me attempt a final (for present purposes) version of the phenomenal property objection.\(^{58}\) Suppose it is claimed that felt pains are physiological items of the \(S_p\) kind. Then the argument from phenomenal property instances goes thus:

(1) Felt pains = physiological items (of the \(S_p\) kind) (where "=" is synthetic and \textit{a posteriori})

(2) If (1), then awareness of felt pain = awareness of physiological items (of the \(S_p\) kind)

(3) If the consequent of (2), then: (a) there is at least one property instance, \(+p^+\), such that "S is aware of \(+p^+\)" entails "S is aware of felt pain"; and (b) there is at least one property instance, \(n\), such that \(n \not\equiv +p^+\), and "S is aware of \(+p^+\) is not entailed by, and does not entail, "S is aware of \(n\)"; and such that "S is aware of \(n\)" entails "S is aware of a physiological item (of the \(S_p\) kind)"
(4) If (3a) and (3b) are true, then each awareness of felt pain will essentially involve awareness of some property instance, such as $+p+$, and no awareness of physiological item (of the $S_p$ kind) will essentially involve awareness of $+p+$.

(5) If awareness of felt pain essentially involves awareness of some property instance such as $+p+$ and no awareness of a physiological item (of the $S_p$ kind) essentially involves awareness of $+p+$, then the former awareness cannot be identical with the latter.

(6) But if awareness of felt pain is not identical with awareness of physiological item (of the $S_p$ kind), then felt pains cannot be identical with physiological items (of the $S_p$ kind) unless the identity were logically necessary.

**THEREFORE**

(7) Felt pains $\neq$ physiological items (of the $S_p$ kind).

I have argued above that awareness of pain does essentially involve awareness (de re) of quality-objects, and that quality-objects are phenomenal property instances. For if awareness of a physiological item (of the $S_p$ kind) is awareness of the quality-object $+p+$, then it would follow that such functionally characterized physiological items give rise to phenomenal property instances. Again, the most that the functionalist is entitled to conclude is: the state of being in pain = the realization of a functional syndrome by physiological items, and that this state causes $+\text{pain}+$, but is not constitutive of $+\text{pain}+$. Thus, (1)-(7) is simply a summary of what has been argued above.
One may think this style of argument can also be used to show that lightning ≠ electrical discharge simply by replacing "felt pain" with "lightning" and "physiological item . . . " by "electrical discharge". But this would not be an appropriate instantiation, for "felt pain" must be replaced by "seen flash". Thus the conclusion would be: seen flash ≠ electrical discharge, although lightning = electrical discharge. Again, felt pain ≠ physiological items (of the S_p kind), although pain (or more fully, being in the state of pain) = being in functional state S_p.

There is an alternative line that I have not considered (not that this is the only alternative there may be), both for reasons of space and (more to the point) because I do not even have a glimmer (or perhaps only a glimmer) of how it addresses the above difficulties. I shall only give an example of this strategy and note that future examination may well be warranted. This way out makes use of a slightly revised version of the brain scanning model of introspective awareness.

Consider David Rosenthal's suggestion:

The most promising account of the consciousness of mental states is to identify a mental state's being conscious with one's being aware of one's being in that mental state and to regard this, in turn, as simply a matter of one's having a roughly contemporaneous thought that one is in that mental state. One can then formulate
a materialist explanation of the consciousness of mental states by hypothesizing that those mental states which are conscious cause one to have the roughly contemporaneous second-order thought and that this causal pattern is a matter of neural or other suitable somatic connections. 

As I have said, I shall not pursue this tack; I only wonder what happens to, when I touch a red-hot iron, the content of my conscious state—do I simply form the belief (de dicto) that I am experiencing pain?

Conclusion

If the arguments adduced thus far are correct, or substantially correct, we are left with a mystery. What are the quality-objects of which I am so often aware? What role (if any) do they play in the life of homo sapiens (or the brutes)? Materialism would have, in a sense, provided the most aesthetically pleasing answer to these and related questions, the most handsome capot with which to cover our ignorance and still our need for definitive answer to that which appears mysterious because of appearances. It is not, it seems to be; materialism has serious difficulties that only one enamored with (unexplicated) physical science (or perhaps anesthetized) can fail to see. What are the alternatives?

Should it be maintained that qualia play a causal role (as unreduced items) in the evolutionary milieu, then some
form of interactoinism raises its banner. That is a bit hard to countenance. Should we take Mackie's tack and view qualia as the by-products of the evolution of neural items, where said neural items (or their physical constituents) form the underlying causal processes responsible for (say) pain-behavior, then we have support for epiphenomenalism. Indeed, Mackie and Wilson (and perhaps Kim) appear close cousins in their views on the (lack of) causal efficacy of qualia—mystery remains. Qualia, as epiphenomenal, may be totally irrelevant to survival. 60

In any event, materialism (M) cannot be sustained. But why should this be considered such a crucial problem for the scientific (naturalistic) world view? While it does, I suppose, leave some sort of door open for mysticism of such-and-such a sort, the failure of materialism does not invite one to become a mystic or believer in fairies. This failure does leave us with a mystery, but surely the scientific edifice is not like the dam—pull out the finger (admit qualia) and the ensuing waters will not drown us. Perhaps we need to further investigate the alleged consequences of the failure of materialism—dualism of some sort need not be indicative of a soft and silly mind, materialism need not be thought the panacea, for there may be no "disease". Still, M is pleasing.
CHAPTER 1--NOTES


2 Haldane and Ross, p. 116.

3 Haldane and Ross, pp. 117-118.

4 Haldane and Ross, p. 222.

5 In what follows I make no attempt to sort out the differences between states, events, processes, etc., and for the most part will use the term "item".


7 By "materialism" I mean the following (see chapter 6):

\[ M = \text{df Every existing item is one that has a property if: (a) the property is either physical, or topic-neutral; or (b) each actual instance of the property is nothing but an instance of a physical or topic-neutral property.} \]


9 Cf. Nagel.


12 Levin, Michael, Metaphysics and the Mind-Body Problem, (Oxford: Clarendon Press, 1979), p. 84-85; see also

287

13 Puccetti, p. 269.


15 I exclude the notion that Mother Nature logically could not have built a successful damage-avoidance system without creating felt pain. This strong claim entails that had felt tickles accompanied damage-avoidance behavior, then felt tickles would be (what we now call) felt pains. But, I maintain, it is conceptually impossible that felt tickles could be felt pains. Thus, the notion that particular felt pains are whatever plays a specified functional role cannot be a logically necessary claim.


17 Rosenberg.


22 Cf. Castañeda, "Consciousness."

24 Rosenberg, pp. 20-21.


27 Nonetheless, as will become obvious, I do not accept this possibility. Following Castañeda, "Consciousness," I accept a postulate of incorrigibility. (See pp. 127-128 in "Consciousness."


30 Smart.


CHAPTER 2—NOTES


3 Block, "Troubles;" "Absent;" Churchland and Churchland, "Functionalism Qualia and Intentionality," mimeo; and Shoemaker, "Inverted."


5 Sober, in "Putting the Function Back Into Functionalism," mimeo, understands functionalism in this way: "... this difference in causal role is physical, and according to functionalism, this wouldn't be enough to establish that there is a psychological difference between the two qualia," (p. 38). Lycan, in "Form, Function, and Feel," Journal of Philosophy, 78, 1 (1981): 24-50, and "Psychological Laws," mimeo, argues this is a phony physical/functional distinction (see chapter 4 below).

6 Cf. Block, "Troubles."

7 Also see Lycan, "Form"; and Shoemaker, "Inverted."

8 In Churchland and Churchland.

9 Again, "+pain+" refers to the phenomenal feature (property instance) in question and is relevant to Churchland and Churchland since they present the inverted qualia difficulty as one of pain/pleasure quale inversion.


12 Descartes (in Haldane and Ross, Vol. II, p. 251) notes that there are three grades of sensation:

(i) Immediate affection of the body; a purely physical process that does not constitute thinking.

(ii) Immediate mental result, perception of pain, thirst, hunger, color, sound, etc.; count as "thinking".

(iii) Judgments about things external, which count as thinking.

I assimilate de re awareness most closely with (ii), a sort of non-judgmental perception.


15 As will become obvious in chapter 6, this is especially relevant to the Adverbial Sensing Theory.

16 Robinson, p. 308.

17 Robinson, pp. 308-309.

18 Robinson, p. 312.


20 Robinson.

21 Churchland and Churchland, p. 2.

22 Puccetti.

23 Puccetti, p. 260.

24 Puccetti, p. 261
Puccetti divides the concept of (physical) pain into three categories. Pain₁ is more or less typical pain most often thought to provoke avoidance behavior. Pain₂ is pain which follows an injury and does not appear to have any preventive function other than immobilizing the sufferer. Pain₃ is rare pathological pain out of all proportion to known tissue damage. As Puccetti points out, concern centers on pain₁; the experience of pain₂ and pain₃ explained as indifferent to the species since it lacks evolutionary advantage to the species as a whole (pain₂) or as the result of neural mechanisms gone awry (pain₃). How plausible all this is, is perhaps debatable; but pain₁ does appear to have evolutionary advantage to the species as a whole. Thus, though there is (perhaps) no biological disadvantage to not feeling pain₂ or pain₃, there is to not feeling pain₁.

If it seems that I have vacillated between these two views, I have. I find the belief that pains+ are causally efficacious a particularly strong belief. Yet I find the violation of accepted scientific principle odious as long as an alternative account can be given which does not violate these principles and which remains consistent with phenomenological data (see chapter 1). Of course, I do not find the belief that pains+ are (mere) beliefs acceptable.

Churchland and Churchland, p. 4.

Churchland and Churchland, p. 8.

Peirce is sensibly sensitive to this; he writes:

The qualities merge into one another. They have no perfect identities, but only likeness, or partial identities. Some of them, as the colours of the musical sounds, form well-understood systems. Probably were our experience of them not so fragmentary, there would be no abrupt demarcation between them at all. Still, each one is what it is in itself without help from the others. They are single but partial determinations.


The difficulty here is aptly brought out in Gunderson's poetical (but poignant) question:
How can I, who am ever at least in part an I or a me and never in all a he or a she or an it to me, be like a him or a her or an it, who or which is always at most a he or a she or an it and never an I or a me to me? That is, since to myself I am not even at least, much less at most, a pattern of behavior(s) and/or neurophysiological events, what on earth could persuade me that someone else's self was nothing but such? But if not that what else could someone's self plausibly be to me?


32 Cornman, *Materialism*.


34 One may be tempted to think this distinction between intuitive appeal and theoretical appeal not to mark a difference worth noting. Construing our commonsense framework as a theoretical structure, we can simply talk of its theoretical appeal as distinct from another theory's appeal. This takes us into deep water and leaves me with the uneasy feeling of floating freely with no place to make port. Which is to say, I do see a distinction, call it what you may, that cannot--not yet--be ignored. The data must be reckoned with, in this instance, eliminated or ignored: qualia are not yet demons--they are, I argue, elements of being.


37 To be fair, they do not use this specific example, though they do make reference to Putnam's "Meaning."
Although Paul Churchland argues elsewhere:

Sensations are just causal middle-men in the process of perception, and one kind will serve as well as another so long as it enjoys the right causal connections. (So far then, in principle they might even be dispensed with, so far as the business of learning and theorizing about the world is concerned. . . .)

In Churchland, *Scientific Realism*, p. 15; cf. Wittgenstein, §§296-307). This cryptic elimination claim has a strange give and take about it: there are sensations but they are not needed for the purpose of learning and theorizing about the world (therefore they are not in the world?). So something in the world (?) remains unexplained (because unexplainable?). I will address a similar claim in detail in the next chapter. (Also compare the discussion on Private Language in Chapter 5.)


Kripke, *Naming*.

CHAPTER 3—NOTES

1 Dennett, Daniel, "On the Absence of Phenomenology," the manuscript from which I quote is dated April, 1974, p. 3.


3 Strictly speaking, for Dennett such folk psychological accounts are not explanations because they do not discharge the task assigned to psychology, viz., answering the question, what is intelligence? Indeed, they assume it since (in brief) a person must have enough intelligence to be able to figure out how to satisfy his desires given his beliefs, etc. Thus, intentional idioms found in the intentional stance are noneliminable. (See Dennett, Brainstorms; and Dennett, "Three Kinds of Intentional Psychology," the manuscript from which I quote is dated May, 1979.

4 In Content, Dennett flirts with functionalism, yet seems to restrict the sub-personal level to the level of realization where talk is of "brains and events in the nervous system" (93ff). By Brainstorms, however, he embraces functionalism where the "flow chart gives us a functional description at what I have called the sub-personal level" (p. 216). He still footnotes, however, Content. Cf. Dennett, "Reply to Stich," mimeo.

5 Dennett claims that what two people have in common when they both believe, for example, that snow is white, is captured by

$$(x) \ (x \ \text{believes that snow is white} \equiv x \ \text{can be predictively attributed the belief that snow is white}).$$

He continues, "[a]ll we need to make an informative answer of this formula is a systematic way of making the attributions alluded to on the right-hand side. . . . Intentional systems are supposed to play a role in the legitimization of mentalistic predicates. . . ." (Brainstorms, p. xvii). For predictive (heuristic) purposes, intentional terms are "legitimized" and have (heuristic) value at the personal level.
Although I prefer to avoid the murky waters surrounding the notion of epistemic access to our own mental states, perhaps the following is an apt notion of privileged access:

It is not logically impossible for someone else ever, justifiably, to deny and correct my sincere word. But it is logically impossible for this to happen very often; i.e., necessarily: in the typical or normal case, or in general, I am the accepted authority on my mental condition.

From Lycan, mimeo, 1974. Pain (+pain+), it would seem, is the crucial case. See below, where this seems to rule out attributing privileged access to the reports of individuals who have been drugged.

Step one is to be found in Dennett, *Content*, pp. 92ff. See also Dennett, "Absence," and *Brainstorms*, chapters 9 and 11.

Again, note that Dennett's question is how do we discriminate pains. Hence, his quest for causal explanation. Here, then, we already get the hint that unless the pain quale figures in a causal explanation, it is gratuitous. See "Posits: For the Purpose Of," in chapter 3 below. Cf. Dennett, *Content*, p. 93.

Dennett, *Content*, p. 93.

Dennett, *Content*, p. 93.

Dennett, *Content*, pp. 94-95.

Dennett, *Content*, p. 95; see chapter 5 below for a detailed examination of this latter passage.
18 Cf. Dennett, Brainstorms, pp. 166-167, where our introspective deliverances are considered theory-laden and mental images "the creatures of a 'posit' . . . exactly analogous to Hume's 'posit' about external bodies." More of this below.

19 I shall consider a variation of this argument below, where Dennett maintains that the concept of pain is inconsistent and thus, no theory of pain is possible.

20 Dennett, Brainstorms, p. x.

21 Dennett, Content, p. 96.

22 Dennett, Content, p. 95.

23 As Stich ("Headaches," Philosophical Books, 2 (1980); I quote from a mimeo) points out in his review of Brainstorms, Dennett appears to flirt with contradiction when he maintains that, strictly speaking, there are no pains even though people do feel pains. Dennett's totally unsatisfactory response is that he is indeed where Stich finds him; "claiming that there are no such things as pains, although of course people do feel pain, and leaving it at that, trusting that the rest of my observations on the subject dissolve the air of paradox," (Dennett, "Reply to Stich," mimeo). For a similar paradox facing Quine, see Lycan, William and Pappas, George, "Quine's Materialism," Philosophia, 6, 1 (1976), pp. 101-130). Cf. Dennett, Content, p. 96.

24 Dennett, Content, p. 96.

25 Dennett, Content, p. 118.


27 Dennett, Brainstorms, pp. 166-167.

28 See Dennett, Brainstorms, chapters 9-11.

29 Cf. Nagel "Bat"; Sellars, "Reply"; and chapter 5 below.

30 Dennett, Brainstorms, p. 225.

31 Dennett, Brainstorms, p. 225.

33 See Dennett, *Brainstorms*, p. 171.


35 Lycan, "Form," p. 42.

36 Cf. Lycan and Pappas.

37 Dennett, *Content*, p. 91.

38 One may be tempted to appeal to the Private Language Argument here and claim that no psychological sentences are reports of private psychological entities. But for a detailed response to this charge, see chapter 5 below.

39 Dennett, *Content*, p. 92.


41 Cf. Sellars, "Reply."


43 Dennett, *Content*, p. 92.


46 As Lycan and Pappas point out, Quinean justification for accepting an existential claim rests on scientific reasons. And, "by "science" Quine means all sorts of explanatory theory-construction, whether highly technical or naive and commonsensical (the sort that all people, even small children, perform" (Lycan and Pappas, p. 123). For an elaborate examination of Quine's views concerning the elimination of mental entities, see Lycan and Pappas.

47 Pappas, pp. 7-8.

48 Again, the question why I want to say certain things is given, at the sub-personal level, an answer via a causal explanation of how I come to utter certain vocables.


Dennett, Brainstorms, pp. 228-229.

Lycan and Pappas, p. 122.

See Pappas,

Lycan and Pappas, p. 123.

See Pappas, p. 9.

Pappas, pp. 9-10.

At a Philosophy Colloquium at The Ohio State University, October 1981. A revised version of Pappas' "Posits" was read.

In Dennett, "Quining Qualia," (the manuscript from which I quote is dated November, 1979), he attempts to argue that the qualia-laden experience and (mere) cognitive events in consciousness (propositional episodes) are not different in kind. He, in effect, argues that our phenomenal experiences are not simples, i.e., unanalyzable, but functionally distinguishable (cf. the experiences of the budding wine taster). Thus,

The depth to which I can penetrate my experience by an analysis is variable, and depends on such factors as training, theoretical knowledge, and the availability of a public vocabulary for marking distinctions (p. 25).

From which he concludes:

But where there is a simple quale at one time, there may be a more or less describable complex at another. The explanation of this variation will not cite intrinsic properties of the relevant events, but functional properties—properties defined in terms of the normal roles and relations of types of internal events to a variety of other functionally defined types or events (p. 25).
Thus, our inclinations and capacities to make discriminations and judgments (e.g., this wine is fat and has too much tannin) cannot be explained by invoking qualia as intrinsic properties. Nor can such invocation "provide suitable referents for the apparent ostensions in those judgments" (p. 25).

I do not follow this argument, for, e.g., wine tasters are only able to make use of public vocabulary to make distinctions amongst their experiences if they share a common experience to analyze. I know, I learned what it is like for a wine to be "fat", but what "that wine is fat" refers to can only be learned via ostension (although science can tell us what causes wine to exhibit that characteristic). Unless you are having (or have had) the experience of tasting a wine which produces that experience labeled "fat", I cannot make you understand what it is like (unless you have had like experiences). You can only learn what it is like, as did I, by becoming aware de re of that experience. In short, there must be a common basis in experience in order for analysis to begin. Compare felt pain: searing, throbbing, shooting, etc. The common basis of wine tasters or pain experiencers is not simply that they share common propositional episodes, or functionally defined properties, for they share these latter only because they share a common experience of which they are aware de re. (Or at least these latter are always accompanied by shared experiences.) Dennett is still attempting to eliminate this ostensive element.
CHAPTER 4--NOTES


2 See Dennett, Brainstorms, chs. 5 and 7.

3 Although perhaps he is, given his reductive identification and a desire to avoid double aspect worries; see below.


5 Dennett, Brainstorms, pp. 123-124; see also pp. 80-81.


7 Dennett, Brainstorms, p. 81.


9 It may be thought that this is merely a pragmatic point; yet, given the relativization of explanation to the purpose and abilities of persons, I think it a highly telling point about the role of explanation in human affairs.


12 Lycan, "Form," p. 34.


16 Margolis, p. 255.

17 Lycan, "Form," p. 32.

18 Dennett, "Three Kinds of Intentional Psychology," the manuscript from which I quote is dated May, 1979).


20 Lycan, "Form," p. 32.

21 Lycan, "Form," p. 32.

22 Dennett, Content, p. 95.


26 Campbell, Keith, "Abstract Particulars and The Philosophy of Mind," mimeo.

27 Lycan, "Form," p. 49; cf. fn. 58 in chapter 3.

28 Lycan, "Form," p. 49.

29 Block, "Absent."


31 Campbell, "Particulars," p. 16.

32 See Campbell, "Particulars."

33 Lycan (in correspondence) wonders who those enamored with qualia "never allow that propositional attitudes (especially urgent desires) have feels."

34 Dennett, Content, p. 64.


37 For a critical examination of Davidson, see Lycan, "Psychological Laws," mimeo.


41 Kim, "Causality," p. 33.

42 In Lycan, "Laws."


44 Kim, "Causality," p. 45.

45 Lycan, "Kripke."


47 Armstrong, Materialist.


49 See Castañeda, "Consciousness."

50 Kim, "Causality," p. 42.

51 Kim, "Causality," p. 43.

52 Mackie argues that we have to recognize property-causes as well as concrete-event causes and thing-causes:
Although it was the one concrete event, Tommy's eating a number of apples, that caused both his immediate stomach ache and his later poisoning, it was the fact that the apples were unripe that caused the stomach ache and the fact that they had been sprayed with paraquat that caused the poisoning.

In Mackie, p. 22.

53 Kim, "Causality," p. 46.

54 Kim, "Causality," p. 46.

55 Kim, "Causality," p. 46; Kim defines the notion of supervenience base thus:

Let M supervene on N, and let P be any property in M such that an object x has P. Then a subset B of N is a supervenience base property ("supervenience base," for short) of the event x's having P just in case B is a minimal set such that x has the properties in B and anything else having these properties has P.

In Kim, "Causality," p. 43.

56 Kim, "Causality," p. 47.

57 Kim, "Causality," p. 47.


60 Mackie, p. 24.

61 Mackie, p. 27.

62 Sober.

63 Cf. Shoemaker, "Inverted."

64 Lycan pointed this out to me.

Cf. Block, "Absent."


Searle, p. 415.

Searle, however, does make an interesting point regarding Dennett's claim that our concept of pain is inconsistent (see ch. 3 above). He states:

From a fact that the set of intuitions we have about a class of objects is inconsistent, and therefore the class is such that none of the objects could satisfy all of the intuitions, it simply does not follow that no such objects exist. Compare: the intuitions we have about chairs do not form a consistent set; therefore, there is no true theory of chairs; therefore chairs do not exist. Or compare: the set of intuitions people have about Jimmy Carter do not form a consistent set; therefore Jimmy Carter does not exist. With the best will in the world there is simply no way you can rescue the formal structure of Dennett's argument that "strictly speaking" pains do not exist. . . .

In Searle, pp. 416-417.
CHAPTER 5—NOTES


2 Cf. Kripke, Wittgenstein.

3 Ned Block distinguishes between Functionalism and Psychofunctionalism by noting that adherents of the former tend to regard functional analyses as analyses of the meaning of mental terms (a priori psychology), while adherents of the latter regard functional analyses as scientific hypotheses (empirical psychology). Lycan identifies himself with the latter.

4 See Dennett, Brainstorms.


8 Feigl, "Crucial," p. 28.

9 Cf. Armstrong, Materialist.

10 Armstrong, Materialist, p. 98.

11 Armstrong, Materialist, p. 97.

12 Wimsatt, "Reductionism, Levels."

13 In what follows I lean heavily on Kripke's Wittgenstein for my understanding of Wittgenstein. I make no claims to being a Wittgenstein scholar.

14 Kripke, Wittgenstein, p. 55.

15 Wittgenstein, p. 81.
Given an understanding of the consequent of (1) as follows:

(a1) There is a difference between someone's obeying and merely thinking that he is obeying the rules for the use of 'E', and

(a2) It is not possible that someone distinguish this difference, i.e., it is not possible that anyone have grounds for determining whether someone correctly or incorrectly thinks that he is obeying these rules.

Moreover, as Cornman points out, although one may think premise (3) dubious because one can follow a rule and not know it, this is not plausible when "obey" is understood in a sense which implies an understanding by someone of what he is following (Cornman, Materialism, p. 88). Thus, "The important distinction here is between 'follow' in the sense of 'obey' and 'follow' in the sense of 'act in accordance with'. For the latter, premise (3) would surely be false, but for 'obey', it seems reasonable to grant it." (Ibid.)
31 With regard to a Robinson Crusoe, Kripke explains:

Does this mean that Robinson Crusoe, isolated on an island, cannot be said to follow any rules, no matter what he does? I do not see that this follows. What does follow is that if we think of Crusoe as following rules, we are taking him into our community and applying our criteria for rule following to him. The falsity of the private model need not mean that a physically isolated individual cannot be said to follow rules; rather that an individual, considered in isolation (whether or not he is physically isolated), cannot be said to do so.

In Kripke, Wittgenstein, p. 110.

32 Wittgenstein, p. 81.

33 Kripke, Wittgenstein, p. 95.


35 Cf. §244 and p. 189 in Wittgenstein.

36 Kripke, Wittgenstein, p. 100.


38 Kripke, Wittgenstein, p. 103.

39 Kripke, Wittgenstein, p. 103. Compare Frank Jackson's case of Fred, who can see one more color than we can. Where we see only reds, Fred sees two colors as different to him as blue and yellow are to us. ("Epiphenomenal Qualia," Philosophical Quarterly, Vol. 32, 127 (1982): 127-136).

40 See §257 in Wittgenstein.
50 Still, that something accompanies our pain-behavior points to an object that stands in need of explanation. Though there may be no fact of the matter relative to meaning, it is a fact that something accompanies my sincere pain-behavior.

51 Cornman, Materialism, p. 92.


55 Castañeda, "Consciousness."

56 I am not sure, but it may be that some such considerations prompt Jackson to maintain:

Considerations grouped under the heading The Private Language Argument are often thought to show that a purely sense-datum language is impossible. But this is a thesis in the philosophy of language which is separate from the ontological thesis I have been urging; it bears on whether our talk about sense-data is conceptually
independent of our talk about physical objects; while my concern has been simply to urge that our talk putatively about sense-data really is about them—be it conceptually independent of physical-object talk or not.


57 Wittgenstein, pp. 88-89; §243.


61 Kripke, Wittgenstein, p. 44.

CHAPTER 6--NOTES


2 Lycan, "Form," p. 31.

3 In Campbell, "Particulars," p. 6; cf. Shoemaker, "Inverted."


6 Jackson, "Qualia," p. 130.


8 Cornman, in Rorty, p. 113.

9 Bernstein, in Rorty, p. 113.

10 Rorty, p. 120.

11 See also Lycan and Pappas.

12 Mackie, p. 20.


14 See Cornman, *Materialism*, pp. 80-81. Cornman takes "is a property" as primitive, and calls nontrivial empirical properties "a posteriori properties", which he defines thus:

\[ \emptyset \text{ is an a posteriori property} = \text{df} \]

(1) \( \emptyset \) is a property; and

(2) For any entity, \( x \), if \( x \) has (lacks) \( \emptyset \), then there is some experiential evidence or theoretical scientific reason sufficient to support a proposition expressed by a sentence of the form "\( y \) has (lacks) \( \emptyset \)". (p. 7).

311


19 Cf. Dennett, Content.

20 Strictly speaking, we should amend (5) by adding (b) (b is a back); but I omit such details in what follows.

21 In Jackson, Perception.

22 In what follows, I tend to take Jackson's line of reasoning, except that I concentrate on sensations whereas he concentrates on after-images.

23 See Kraut, Robert, "Sensory States and Sensory Objects," Nous, 16 (1982): 277-93; I use Kraut's examples, but change the indexing.

24 Kraut, pp. 278-279.

25 Kraut, p. 289.

26 Kraut, p. 282.

27 Kraut, p. 292.

28 Kraut, p. 292.

29 Kraut, p. 287.

30 Kraut, p. 282.


34 Wilson, p. 59.

35 Wilson, p. 59.

36 In Wilson, p. 62.


38 See Cornman, Materialism, p. 9. This is not exactly as Cornman renders materialism, but it captures his intent, I believe. Also, see footnote 14, this chapter.

39 Wilson, p. 62.

40 Cf. Cornman, Materialism.

41 The same criticism I invoke here applies, mutatis mutandis, to Wilson's other examples, namely, atmospheric halos, coronas, and after-glow.

42 Wilson, p. 63.

43 Wilson, p. 63.


45 Wilson, p. 63.

46 Also see Cornman, Materialism.

47 In the case of a dog, say, we can track down the neural correlate of pain (directly) via mere behavior. There is still, however, an indirect link with (human) pain reports. Of course the dog cannot tell us we are mistaken; a human can. And if a human tells us that, despite his behavior, he did (not) feel pain, we accept his report. Jones may wince and cry out, not because he is in pain, but because he remembers he forgot to lock the safe. Not all pains are accompanied by overt behavior, and when we "look inside" all we can correlate the inner goings-on with, is
verbal reports. Now, why think that (in the case of humans) overt behavior is pain-behavior? Again, Jones. We do so because of what others (ourselves) say (feel); only first-person statements about mental states reflect reliance on our introspective awareness of how they feel. That mere behavior serves to track down the neural correlate of pain in a dog, itself indirectly depends on our having already identified certain behavior as pain-behavior via (human) reports.

49 Block, Readings, p. 179.
50 This version is taken from Lycan, "Topic Neutrality and Bradley's Objections," mimeo. Much of what follows was generated as an attempt to meet objections he raises.
51 See Block, Readings.
53 Cf. Smart, "Sensations."
56 Block, Readings, p. 21.
57 Mackie, p. 21.
58 I take this from a version attributed to Alan Hausman, though he may well not accept this particular rendition.
60 See Jackson, "Qualia."
BIBLIOGRAPHY


----------. "Are Absent Qualia Possible?" *Philosophical Review*, 89 (1980).


Churchland and Churchland. "Functionalism Qualia and Intentionality." Mimeo.


-----— —. "On the Absence of Phenomenology." The manuscript from which I quote is dated April, 1974.


-----— —. "Three Kinds of Intentional Psychology." The manuscript from which I quote is dated May, 1979.

-----— —. "Quining Qualia." The manuscript from which I quote is dated November, 1979.

-----— —. "Reply to Stich." Mimeo.


---------. "Psychological Laws." Mimeo.


---------. "What is it Like to be a Bat?" Journal of Philosophy, 78, 1 (1981).


