107-441B Keith Douglas
Prof. Bunge 9630153

The Putative Philosophical Relevance of Neuroscience

Many philosophers and scientists in recent years have adopted various versions of the thesis that neuroscience is directly philosophically relevant. (A short list would include: Bunge 1980; Bunge and Ardila 1987; P. M. Churchland 1995; P. S. Churchland and Sejnowski 1996; P. S. Churchland and P. M. Churchland 1998, Gold and Stoljar 1999; Dennett 1991; Hebb 1980; Ramsey, Stich and Rumelhart 1991). For views somewhat[1] in the contrary position in various forms, see Searle 1992; Penrose 1989, 1994, 1997; Jackson 1982; Nagel 1974) Despite my view that “things scientific”[2] are useful in philosophy, I feel that some of the recent enthusiasm is either misplaced or confused. In the present paper I shall sketch what I call a “minimalist” viewpoint on the relevance of neuroscience to philosophy. Throughout, I will concentrate on how my view differs from that of the others. Despite my generally critical tone, it must be said that for the most part I do agree with the general thesis that findings and methodology in neuroscience have important philosophical implications. I only quibble over details in order to better sort out the morals to be drawn.

This paper will consists of two broad parts each with subdivisions. One of the broad parts will consist of the sketch of the minimalist viewpoint, and the second one shall briefly respond to possible criticisms of the viewpoint sketched. In both sections viewpoints of other philosophers and scientists will be brought up as necessary. The present paper will presuppose a basic familiarity with neuroscience generally and make reference to literature in this subject as well to the philosophical literature.

Section 1 - Minimalism

I see no reason to define “minimalism” explicitly. Instead, the view sketched will constitute something of an implicit definition of the term. (I only coin the term so that it may be referred to in contrast with the views of others who generally speaking have adopted names for themselves.)

---

[1] All of these thinkers agree that it is relevant as well, just mainly in a negative sense. They think the relevance of neuroscience to philosophy has been “overstated” or misrepresented.

[2] This should be taken as meaning methods, findings, philosophical presuppositions, theories, tools, and so on.

In this section, I shall present 4 distinct areas of philosophy in which neuroscience is relevant. These will be much the same as others who also feel that neuroscience is relevant to philosophy, however, our differences will lie primarily in the details. The areas to be presented are: “the mind/body problem”, “ontology”, “epistemology” and “philosophy of perception”. There are potentially others (some others which have been suggested in the literature particularly often: “philosophy of science”, “philosophy of mathematics”, “ethics”, and “semantics”); I shall ignore them in the interests of making the present work manageable. I also do not mean to suggest that these areas are mutually exclusive. As we shall see, they overlap quite strongly as they do in a “traditional” philosophical system. On to the first one, then.

Section 1a - “The mind/body problem”

This section is placed first amongst the areas to be explored because it is the area in which neuroscience has the greatest promise and in fact where it is used most directly by most philosophers.

In my view, neuroscience completely vindicates the thesis that “minds” are not a thing but a process. This discovery, however, is not narrow enough to make the distinction between several positions in the philosophy of mind, as several positions are said to be compatible with neuroscience. Certain kinds of functionalism (particularly so called “homuncular” functionalism) are said to be such (Dennett 1991); others say that neuroscience supports emergent materialism (Bunge 1980; Bunge and Ardila 1987) and still others say that neuroscience supports eliminative materialism. (P. M. Churchland 1995) My view is that one should “split the difference” amongst these viewpoints; as we shall see, the differences between them are very slight.

This conclusion arises from several considerations. The first is it that is very difficult to know what “mentalistic” predicates are capable of being translated into neuroscientific terms. It would be very strange indeed, though certainly not an unheard of phenomenon in the history of science[3], if none of our prescientific categories were capable of being “mapped on” to what is discovered by neuroscience. The nature of this “mapping on” is a bit vague, and is part of the point at issue. The thesis that none are capable in the strict sense of being so mapped is the view of the eliminative materialist (were there any strict eliminativists left). On

[3] One possible example might be “energy” and how it turned out that energy is not a thing but a property. Traditional viewpoints from around the world share this feature and it still infects the natural language of scientific cultures such as our own.

the other hand, the emergent materialist is more inclined to be enthusiastic in the other direction. This position also suffers from an additional caveat. Since mentalistic terminology is very vague and imprecise, it is difficult to tell when one has found a (terminological) correspondence. An example will be helpful here.

Take the concept of “memory”. Psychologists recognize that there are numerous “memories”. Some of these have been somewhat localized in the brain (see Squire and Zola-Morgan 1991). However, since prescientific talk of “memory” is that it is unified (or at most divided into two or three “units”), in what sense has memory been localized? Simply counting the units will not help as the naive conception of what they do might be wrong as well. A theory of reference (such as that in Bunge 1974) will not help either, as has been said already, these prescientific conceptions are quite vague.

I think the solution here is in fact to first exactify at a more broadly psychological level what is meant by learning, motivation, attention, memory, and so on before postulating or announcing the discovery of the relevant neuronal systems. Otherwise, there will no doubt be charges of the fallacy of redefinition. (As we shall see later, some of the critics of neuroscience’s influence on philosophy are going to say that the whole project is guilty of this. I do not share their pessimism, but it is a legitimate concern in a certain limited way.) Thus whether one should be an eliminativist or an emergentist is done for each “folk psychological” concept individually. I think it is safe to say that “consciousness” at least in a “cartesian materialist”[4] sense has been eliminated[5]. On the other hand, memory seems to be about half way to being eliminated (mainly because it is so much of a function of the brain (or at least the cerebral cortex) as whole. See, in particular, Beardsley 1997.), and I suspect that “motivation” will be more vindicated. But these are predictions based on the current state of conventional cognitive psychology and cognitive neuroscience. Since there has been little or no attempt to perform the preneuroscientific exactification I mentioned, these predictions are little more than opinions.

[4] A cartesian materialist is someone who thinks that there is a privileged place in the brain where all “conscious” experience “gets presented” to a “self” or something like one.

[5] The view that it is to be identified with some of the functioning of the frontal lobes (as in Bunge 1980, for instance) of the brain is somewhat contentious. See Milner and Petrides 1984 as well as Beardsley 1997. Nevertheless, “frontal lobes” is broad enough to support the main point, as it does contain many distinct anatomical subsystems.

The second way in which the "split the difference" thesis should be useful is in a reconciliation between the functionalists and the other two positions. Now, it is of course true that some functionalists are not complete materialists (see Fodor 1980 and the next section for details), most are, so there should be some way to split the difference here, too[6]. How so? I propose that postulating that "mental processes are brain processes" and that postulating that "mental processes are only brain processes", where brain means an appropriate neural system made of the sorts of neurotransmitters, neurons, and glial cells as found in animals are in fact two different claims that should be distinguished. I am perfectly in agreement with the thesis that all known mental processes are processes of these kinds of brains, but that does not rule out *a priori* other kinds of brains[7]. This is the point I am suggesting one should concede to the functionalist. It has been suggested by some (notably P. S. Churchland & Sejnowski 1996), that what matters in brains is there connectivity (number and kinds of connections) as well as the methods of synaptic weight adjustments. If this is true, this would suggest a weak form of functionalism is correct.

Section 1b - "Ontology"

As has already been remarked, there is some consensus that neuroscience does support an ontology of materialism. Here I am in complete agreement with the viewpoints on this subject held by many. In the "minimalist" account of neuroscience and philosophy I am sketching, I adopt the viewpoint that neuroscience does provide evidence for materialism[8], in spite of several popular objections. I shall explore several which are relatively recent and leave the discussion of the older versions of dualism/idealism to the literature (Bunge 1980, Hamlyn 1984, Dennett 1991). I am doing this here rather than in section two as I feel that the mistakes are instructive ones for any discussion of ontology.

The first of these arguments centers around the objection to dualism

[6] More on this in the next section, but I do not feel it is possible to split the difference with the dualists and idealists. We shall see two way that have been proposed in recent years to do this and show that they are unsatisfactory.

[7] Alternatively, if one wants to phrase this thesis differently: there could be other mentating systems. I regard these two phrases as synonyms.

[8] I am also of the opinion that neuroscientific research actually presupposes materialism, but I will briefly discuss this in the second section (on objections) of this paper.

that uses the law of the conservation of energy for support. The argument runs something like this. The law of conservation of energy can be violated, according to quantum mechanics, by an amount ∂E if the time "by which it must be payed back" (∂t) have a product less than the normalized Planck's constant. Hence the interaction between the body and the immaterial mind has "time to interact".

There are three problems with this, irrespective of the traditional problems (interactionism, etc.) with dualism. The first is that it ignores the objection being made - namely that conservation of energy is observed in the brain's functioning. (The critic will answer that this amount is going to be so small that and spread over such a large time that it will be difficult to detect. But of course this renders the account even more *ad hoc* than it was to begin with.) The second is that some (Bunge 1973) have suggested that this inequality is in fact not legitimately part of the quantum theory at all, and so this form of dualism is explaining the obscure by the (almost) equally obscure. Third, it commits the level skipping error - it would be very strange if things at the biological level could "make use of this" irrespective of the underlying chemical level. In this sense, it is antievolutionary as well. It is also antievolutionary in another respect as it proposes a substantial discontinuity between human brains and those of other animals.

The second of the objections is a bit more principled, and does not require any "spooky physics" to get off the ground. This family of objections (Nagel 1974; Jackson 1982 and others) concerns "private experience". The claim goes that any materialist account of "how the mind works" necessarily leaves out the "what it is like to be" whatever that has the mind. (This is using Nagel's language). This is also sometimes expressed in terms of materialism/science/etc[9]. cannot do justice to "points of view". The claim is usually expressed in terms of something like: because science necessarily studies from the third person perspective, and the very essence of mentality is the first person perspective on things. Hence because science is necessarily incomplete in an important respect, even if one knows everything about the brain, there must be something more than what neuroscience discovers - the first person perspective. This argument is very seductive, however, it does not establish the conclusion it wants. Let us see why.

---

[9] To the objectors credit, they recognize that science presupposes an ontology of materialism.

Neuroscience, it is true, works from the third person perspective. However, with that in mind (or should one say in brain?), it is simply a *non sequitur* to claim that there is something missing when that approach is taken. How does one make the claim, without begging the question against a (or several) neuroscientific account(s) of mind, that something will be left out when “neuroscience is completed”. I do not know of any “complete” sciences. Hence I have no intuitions either way about what it would be like to have a complete (whatever that would mean) scientific picture of the brain. Perhaps from that perspective (if I had it, say, of the reader’s brain) I <u>would</u> know “what it is like to be” the reader. Nagel (as we have been looking primarily at his expounding of this viewpoint) cannot insist that something is left out without begging the question against the materialist.

Another ontological area in which neuroscience has been useful, which is underappreciated, concerns the principle of lawfulness. Pseudoscientific accounts (for instance psychoanalysis) of the mental often reject the principle of lawfulness when it comes to the “functioning of the mind”. (See Douglas 1998b). Neuroscience firmly grounds the research into the mental in states, processes and things that are well established to be lawful. (In other words, the postulate of the principle of lawfulness is well established for neurons and so forth.) Of course, this does cause some backlash. For some reason, some (such as found in the proceedings of McGill University course 107-590A, fall 1998 semester, and echoed in [for instance], Taylor 1988) have thought that asserting that humans are “subject to” neurological laws meant “giving up freedom.” This conclusion was (of course) regarded as undesirable, and hence the idea of the principle of lawfulness, at least as it pertained to ‘mentality’ was rejected. (Sometimes materialism was given up for similar reasons.) This confusion is relatively wide spread, and it behooves any philosopher concerned with neuroscience to sort it out. (The concern over human freedom is of course one of the recurring themes in philosophy, particularly centered around scientific discovery. It is therefore not at all surprising that neuroscience has had this effect as well.)

What should “neurophilosophers” and neuroscientists do to allay fears in this area, without giving up an ontological principle so important in scientific research (Bunge 1977)? It suffices for the most part to remark that admitting that there are objective regularities says nothing about the origin or cause of the objective regularities. Freedom (as understood as involving partially internal causes) is thus perfectly compatible with lawfulness. One should also stress that lawlessness would involve less freedom, not more, as one could not know that one could even successfully move one’s pinky finger all the time in a lawless universe.

This brings us to the penultimate point in our discussion of the minimalist account in the domain of ontology. This concerns the nature of cause. Here neuroscience should remind the philosopher of the multiplicity of the species of causation, an area which is very underdeveloped. Philosophers generally admit that causation is not all of the “billiard ball” type associated with Newtonian mechanics. Neuroscience tells us of at least three other kinds. Firstly, it tells us of causation at the chemical level. Neurotransmitters act at the chemical level of a cell, causing and inhibiting chemical reactions in the cell they influence. This causation does not occur in anything like the ‘mechanical’ way. It also tells us of causation at the tissue or organ level, as when a muscle is contracted and relaxed repeatedly by neuronal activity, one is able to walk, talk or have one’s heart continue pumping blood. This yields the final sort of causation that neuroscience reminds us of - the personal or psychological level of causation. By means of all these underlying levels, we can throw a ball or recite a poem, and influence causally the systems around us, including other people. A final note each of these varieties may also be involved in determination[10] of sorts that is not strictly causal. In other words, neuroscience should remind philosophers of the existence of stochastic processes, in particular, as it appears that to some degree neuronal firing is such a process.

This gives us our final point in the ontology involved in the minimalist account. As we noted above, there are several kinds of causation a philosopher should keep track of and that neuroscience reminds her of[11]. Each of these occurs at a different <u>level</u> of reality. Neuroscience reminds us that there are several levels of reality, each ontologically distinct but not separate from the others. We should especially take note of the differences between the neurological level and the physical level. A failure to appreciate these differences is perhaps is the source of some of the confusion present in Nagel’s (1974) work, discussed above. Whether one wants

[10] Note that this appears to dissolve the age-old free will versus determinism debate, but I do not have time to explore that issue now. It does seem to be another area of philosophy which may be greatly aided by neuroscientific findings.

[11] There are of course other kinds of causation that neuroscience need not remind us of, but the very fact that it does remind of us of several might help to “jog our memory” for others.

to be an eliminativist, an emergentist, or somewhere halfway between (see above in the section discussing the mind-body problem), one ought to recognize that the neurological level has emergent laws that are not found at the physical level (or the chemical level) underlying it. Nagel’s insistence that science ignores points of view is true of physics, however, it is not true of neuroscience. Admittedly, neuroscience does not directly use the first person perspective (as we saw above), but that doesn’t mean it isn’t studied. <u>Cognitive</u> neuroscience is precisely about this issue. (Another related lesson to philosophers here is to avoid physics-worship. Physics is the most successful of the branches of science, so it is not surprising that many philosophers (and indeed, many scientists) want to “ape” it. This is an ontological lesson in part, as it is meant to remind us that there are different things at different levels. It is also an epistemological lesson, of which we shall see more anon.)

<u>Section 1c - “Epistemology”</u>

Here we have two parallel lessons to present as part of the minimalism. The first concerns what might be called the outside-approach, the second the inside-approach. The first is called such because it concerns how neuroscience informs us about our own acquisition of knowledge, which is relevant to epistemology. But it also informs us by examining how traditional and modern epistemologies fare in the investigation <u>within</u> neuroscience. The latter overlaps strongly with issues in the philosophy of science generally speaking, and we shall see a bit more about it in this light later.

What does neuroscience tell us about the acquisition of knowledge? A great deal, and with much promise for the future, as it appears at long last we know something of the mechanisms involved. This promising start into how knowledge is acquired by subjects does not mean the road ahead is not smooth. Several principle controversies in the light of the success of neuroscience arise in this area; I shall examine four. The first concerns the importance of traditional cognitive psychology, the second the relevance of technological disciplines such as artificial intelligence. The third, which is the combination of the previous two, and further related to the mind-body problem is the so called “computational theory of mind” and the most misunderstood (by both proponents and critics alike) interesting (to the present author at least) in terms of knowledge acquisition. The fourth area of concern I will look at concerns acquisition of languages. (These are not meant to be mutually exclusive problems. How one gets resolved in one’s philosophy of neuroscience influences how the others do as well.)

Traditional cognitive psychology involves both behaviourism à la Skinner, as well as the discipline that studies the “phenomena” of cognition. (See Bunge 1983 for an explanation of this latter characterization.) The former is quite rightly regarded as being a dead discipline. No cognitive scientist, whether neuroscientist, psychologist or artificial intelligence researcher these days is inclined to ignore the inner state of the organism or machine they are studying (Medin and Ross 1997). Neuroscience is in fact very promising here, as it allows us to better know about the inner states shunned by the behaviourists. Nevertheless, there is some controversy over how much the actual “implementation details” (i.e., to what extent and in which way, is neuroscience important) matter.

Let us look at an example. Studies concerning priming indicate that the internal state of an organism is certainly relevant to their behaviour on future occasions. We can discover facts here independently of the basis (neurology in our case) of cognition. Neurology provides a constraint on this research, however. If the psychological explanation one has developed seems to require postulating neurologically dubious structures, we should perhaps rethink our the conclusions of the psychology. (This is a special case of the more general epistemological principle of level agreement.) However, it is also important to note that neurological investigation alone equally proceeds in the dark. Investigating any system requires both knowledge of structure and of function - a purely neurological investigation will tend to obscure the latter. (This is recognized even by the eliminative materialists, if only because without the cognitive level being looked at the strong elminativist will not know what to eliminate!)

The upshot of this is that traditional cognitive psychology does have a future in telling us about the acquisition of knowledge, but the degree of its importance ought not to be overstated[12]. The same goes, *mutatis mutandis*, for neuroscience itself.

Next, let us now look at another question in this area to consider. What is the relevance of artificial intelligence to the study of knowledge? Does neuroscience suggest that the enterprise of artificial intelligence is doomed to fail as some have suggested? It is true that constructing

[12] I find that the “symbiosis” or “cooperation” pair fascinating - further work on such “mutually reinforcing” pairs of sciences is indicated. (I know of no other pairs which are “symbiotic” to this extent - <u>perhaps</u> biochemistry and genetics.)

artificially intelligent devices presupposes some knowledge of existing intelligent things (Bunge 1983). But in what respect? What must be known about them? The important question here is whether neuroscience tells us that only (certain) neuronal systems can learn. Some thinkers have postulated this very thesis and consider this as an affront to certain kinds of artificial intelligence research. While there is insufficient time for the present work to refute all the biological objections to artificial intelligence (for some work in that direction, see Douglas 1998a) it is important to note that (e.g.) Bunge's (1980, 1983) postulate is exactly that, a postulate. Hence it may very well turn out to be false; it is hence no <u>argument</u> against AI to say that only systems with plastic neural assemblies can learn. It does not seem plausible to suggest that neuroscience has informed us of this. It has informed us somewhat of how neural systems learn and says very little about learning systems in general (except, perhaps, that they are remarkably complex things and that simple systems clearly do not learn). This viewpoint does not mean any of the current approaches to AI are correct or fruitful either.

Note also that a family of the earlier failures in AI does tell us something about human acquisition of knowledge even with its lack of contact with neuroscience. (I do agree that it failed precisely because of a lack of contact with cognitive psychology and probably neuroscience, but that is another story for another time.) This failure is of what Hofstadter (1985) calls the "Boolean Dream" - the idea that one could create an AI by piling rules with metarules on top with metametarules on top of them, and so forth. This is a strictly negative result - it has told us that our knowledge systems do NOT work like this.

Finally, there is at least one open problem in AI does pose a question that is interesting for epistemologists to consider. As "promised", it is also somewhat independent of neurological functioning. (It is of course very plausible that a further study into neuroscience will in fact yield the solution to the problem, but the problem would possibly not have been posed without research in AI.) This is the notorious **frame problem.** (For more and sketches of a solution, see Dennett 1998).

A brief recapitulation of the frame problem follows in order to bring the reader up to speed. Basically the problem can be stated as follows: how does a particular cognitive system (in the case of a biological creature, the creature's brain) know what inferences to draw in a given situation. Too few, and the creature will have insufficient information about the environment to act; too many and the creature will be overwhelmed by the process and will be unable to act until all these (a potentially infinite number) of conclusions are drawn. Two important features to note of this problem are as follows. Firstly, that the problem is not committed to the idea that these inferences are done **consciously** in any sense. In other words, we do not have to imagine a squirrel sitting at the base of a tree with modus ponens flashing through its head - these inferences can be “prelogical”. Animals do draw inferences in other ways, however[13]. Secondly, it is important to note that generally a creature would not know the consequent of drawing a conclusion until it is drawn - that IS the whole point of inferring in the first place. Hence it cannot know what ones to draw until they are drawn[14].

An obvious, but insufficient, biologically oriented answer to the frame problem concerns evolutionary and developmental biology, including evolutionary and developmental neuroscience. But simply saying those animals who drew the correct influences and the “right” number[15] of inferences were the ones who survived and got selected for does not fully answer the question. For one, it does not specify the mechanism by which the neural mechanisms “worked partially” did there thing, which is the interesting point here. Secondly, it does not allow for animals (particularly humans) to draw inferences in domains where the role of selection is unlikely to have played any part. (For instance, art, some aspects of society, science, technology, and so forth.)

My goal here is not to present any solution to the frame problem, but to make the reader aware of a general question of epistemology prompted by AI research that is independent of neuroscience for its posing. Alternatively, it emphasizes the relevance of other disciplines, even technologies, to the study of brain functioning. (As remarked earlier, the

[13] Actually, in some sense, so do plants, protists and monerans. A plant will “draw the conclusion” that a strong light placed near it is the sun, an amoeba that a certain kind of chemical gradient in the water indicates a source of food, and so forth. Generally, though, the more complex an organism is, the more difficult their frame problem is. However, I do not think the frame problem “affects” any non-vertebrates, except for (perhaps) cephalopods, as they have relatively “advanced” nervous systems.

[14] There have been claims that the frame problem even refutes materialism. It does not, if only for the reason that any account using “immaterial things” would suffer from the same problem - or at least, could not be postulated not to without begging the question.

[15] Note that this number of correct inferences, like virtually everything in an evolutionary problem, is context dependent.

problem may not be so independent of neuroscience for its resolutions(s).)

Our third area of investigation concerns whether or not the so called "computational theory of mind" (CTM) has been discredited by investigation in neuroscience. In other words, is there any hope left for the CTM to tell us anything about the acquisition of knowledge? (Bunge 1983, for instance, seems to contain the viewpoint that neuroscience has discredited 'it'.) This way of putting it is a bit misleading, as it presupposes there is one "theory" in this area[16]. Some versions of this thesis do not make any contact with neuroscience at all, and have yet to produce much in the way of discoveries concerning knowledge acquisition (for instance, that of Fodor 1980). The latter is sometimes called "Turing machine functionalism", though it is disputed whether Fodor actually ever held it in that strong a form[17]. Some slightly "milder" versions follow the tradition of the traditional cognitive psychologists and feel that they (those adopting their version of "computationalism") will "meet the neuroscientists tunnelling through from the other side" (Pinker 1997). A sort of middle position is adopted by Dennett (1991) where parts of the "computationalism" does make contact with neuroscience. Finally, we have the neurocomputationalism of the Churchlands, who make the case in a very different way for a form of "computationalism" and also rely heavily on contemporary neuroscience. I will survey these viewpoints and show that each has lessons that are complementary to and yet related to the neuroscientific project.

The "Pinker" form of computationalism makes a heuristic point that is perhaps overlooked in neuroscience's influence on philosophy. Neuroscience is often felt to be "removed" from every day experience. Even the very best thinkers need some prodding to see how intelligence can arise out of nonintelligent things. It is difficult to show how to do this within the framework of neuroscience proper. Pinker (1997) uses an example involving Turing machines in order to make this "smart-stupid" point. This does <u>not</u> commit him to the crazy thesis that the brain is a Turing machine. (But note

---

[16] I agree with the critics who point out that there is no theory (i.e. hypothetico-deductive system) proper to be found in any of these versions. Nevertheless, I shall investigate what they have to offer vis-a-vis neuroscience and the acquisition of knowledge.

[17] The point of contention is over whether anyone (and Fodor in specific) ever actually held the thesis that the brain literally instantiates a Turing machine. (Note that this statement of the putative instantiation is a bit ambiguous between a Turing machine as a paper tape and sensor type mechanism, and that of a more conventional computer CPU or the like. This ambiguity fuels the debate to some extent. Obviously Fodor never claimed there is <u>literally</u> a paper tape in one's brain.

that rejecting Turing machine functionalism (which is a position that perhaps nobody really held (Dennett 1991)) does not mean rejecting computationalism. What exactly Pinker is committed to computationally is still (to me at least) unclear.

On the other hand, Dennett's computationalism is much better spelled out. For example, his computationalism makes use of the serial/parallel distinction in computer science. (He does not, *pace* Churchland 1995, use this to claim that consciousness is "serial" in nature.) Does this inform us any more about knowledge acquisition independently of neuroscience? In a curious way (not discussed much in his book, unfortunately), yes. Medin and Ross (1997) make use of this distinction (independently of Dennett) in their discussion of certain discoveries in cognitive psychology. Facts such as minimal attachment in language parsing only make sense in terms of a serial "processor" and some others, such as the Stroop effect[18], only make sense in terms of a parallel one.

The most striking and unusual of the computational approaches is that held by the Churchlands, however, as they use it as a **guiding principle** in their use of neuroscience[19]. Despite its strong ties (justified or otherwise - I shall return to this in due course) to "computationalism", it is important to note that it is radically different than most traditional accounts of this name. This is the case in three important respects. Firstly, the Churchlands realize[20] that asserting that the brain is computer-like (in some respect) is a matter of **stipulation** (we shall return to this when I discuss the merits of their proposals). Secondly, they are quick to point out that when they say "computer-like", they do not mean serial, digital, von Neumann computers. They are quite correct in pointing out that in principle there are computers which are parallel (no problem there), analog (a bit unusual) and not von Neumann in organization (strange)[21].

[18] Appendix A to this paper contains a brief discussion of this effect as it is relevant to several parts of this paper.

[19] It is important to stress that this thesis is not the crazy thesis of only a pair of philosophers. At the very least, it is the crazy thesis of a bunch of neuroscientists as well. See (e.g.) Nadel *et. al.* 1994; Koch and Segev 1998.

[20] Explaining this thesis is quite important, as I agree with it. However, I do not feel that the present paper would be at all manageable with this digression in it. I remark instead that at the very least the success of the computational neuroscience program is far ranging, as we shall see. At the very least, this thesis has been an extraordinarily useful heuristic.

[21] This thesis is strange as virtually all computers ever made (past and present) are von

Thirdly, they also give a novel account of computationalism that is radically different from most accounts argued against by those who are not sympathetic to “computationalism”.

What are the merits of the computationalist proposal then? How does this thesis inform neuroscience? First, it is important to realize what exactly is meant by a computation here. This is the point of alleged insight and the greatest point of contention. The claim is that the nervous system computes (and this is far below any “conscious level”) because it takes as input (either from other nervous system structures or directly from the environment) an “activation vector” which gets transformed by a given nervous system structure into another kind of activation vector as both a ‘function’ of the input and the internal state of the system, whose effects on other systems of the body produces behaviour and responses of all kinds. No algorithm[22] of any kind is said to be used in the usual contentious sense; instead a transformation occurs. This computational approach is said to be useful because it tells us why (for instance) a crab is able to “figure out” that its pincers have to be positioned at such and such an angle to its body to grab a piece of food, and so on. This angle has to be “produced” by the nervous system somehow, and the suggestion is that without talk of it being computed, there would be no principled way to talk about what goes on inside its nervous system. The claim is that one would be stuck with vague language such as the “produced” I used earlier. The same goes for any other example of creatures performing tasks.

In this sense, the computational approach acts as a constraint on the neurological investigations, because it tells us what the parts must “do”. In the crab example, the outputs from the motor cortex have to move the pincers into the right place. It is vital to see, however, that the Churchlands are not claiming that there is somehow encoded in the brain something like a program written in a procedural computer programming language. Thus the first lesson to be learned (if it is legitimate) from this version of computationalism is the constraint approach. The computationalism acts as a constrain by telling the neuroscientist to look for well specified inputs and outputs as well as inner states and transformations. It also acts as a time constraint; we know from the study

Neumann machines.

[22] The so called “learning algorithms” in the context of connectionist models are meant to <u>model</u> a process that goes on in genuine nervous systems. Whether these do or not is another story, and unrelated to the use of algorithm here. The learning algorithm is, so to speak, “embodied” in the cell assemblies in a nervous system.

of algorithms in computer science that certain ways of processing would take too long to accomplish and hence would conflict with observed times of reaction, and so forth. Therefore, what the organism is doing must be using a process of such and such an efficiency.

The second potential merit of Churchlandian computationalism is that also suggests how to design artificial models of certain aspects of nervous systems. This is not just an idle promise of an unfulfilled research program. Many of these models have been constructed, some with remarkably biological features. For instance, nobody knows any algorithm that performs the operations over “least vision parts”[23] that a owl retina does. Yet researchers have constructed an artificial neural network (see Churchland 1995) that is remarkably like this. This means that it is similar in outputs given a similar internal state[24] and similar inputs.

Epistemologically, neural networks are interesting because they “learn” without being taught specific rules, unlike the systems of “traditional AI”[25]. They thus have more biological plausibility. (Very few researchers think these days that “rules all the way down” is a good model for human mental functioning, and, *a fortori*, machine functioning, should it prove possible.)

This is not the place to survey the successes and failures of artificial neural networks taken generally. However it is important to realize they are excellent for checking understanding of at least the gross features of neural assemblies. They thus teach us that perhaps the “substance independence” wanted by the functionalists (see above) is

[23] Or in other words, the minimum discriminated “pieces” of visual experience. (All vision is ultimately discrete, at least spatially, as there are a finite number of retinal cells in a given area with gaps between them. See, e.g. Sekuler and Blake 1994 for details.)

[24] I stress the “internal state” here, as many critics of “computationalism” have correctly pointed out that internal states are often overlooked in some versions of “computationalism” and/or AI.

[25] I put “traditional” in quotation marks, as some (particularly Douglas Hofstadter) working in AI and “the” (as we have seen there are really several) computational theory of mind have recognized for at least 20 years that this is not plausible biologically or computationally. It is prohibitively computationally expensive, which suggests here’s another place where neuroscience can learn from computationalism again. If your neuroscience postulates that the brain requires some serial rule follower underlying everything of a certain kind, one should think again, because computer science tells us that such a “processor” would be grossly inefficient and would almost certainly get selected against.

correct. This is a tentative conclusion, of course, and larger scale networks must be constructed before one should seriously entertain this possibility. (We shall return to possible objections in section two of this paper.)

Our lesson for minimalism from the versions of computationalism we have seen is two fold. One is that the versions of computationalism vary tremendously in biological plausibility - from none (Fodor’s immaterial Turing machine) to very much plausible - the computational neuroscientific approach of the Churchlands. Second is that they all give different answers concerning how knowledge works (e.g. how knowledge is constructed in the brain, and how it relates to the external and internal environments of the organism, etc.). The computational neuroscience of the Churchlands is very promising when it comes to learning and knowledge. It has already shown (Churchland and Sejnowski 1992) that some learning follows the Hebb rule (Hebb 1949) and some does not, and further has refined the Hebb rule into subcategories. I feel at this stage that drawing any greater conclusions for this branch of epistemology is premature, as not enough very large scale and multi-use artificial neural networks have been constructed. The ones that do learn interesting things (for example how to recognize faces, how to speak English words) are very unifaceted. Hence any general epistemological lessons should be held off until the networks are more general. As, after all, mammalian nervous systems are capable of doing many different tasks, often simultaneously. Another final lesson concerns an epistemological issue in technology. That is, consideration of neurology has led us to the conclusion (if we accept the Churchlandian view) that “computer” is in some sense a conventional designation - i.e. what it picks out is somewhat conventional.

Let us now look at how neuroscience affects the epistemology of language acquisition. The most important result here is the apparent falsification of the “language organ” thesis of Chomsky. Individuals who are born with their language center in one hemisphere of the brain, and then suffer childhood brain injury very often “produce” another language center on the opposite side of their brain (Rasmussen and Milner 1975). Multilingual adults often lose language selectively upon brain damage. These findings suggest (do not conclusively demonstrate, however) that there is not a single language organ in humans. (A suggestion that the language organ be identified with the Broca and Wernicke’s area of a specific hemisphere is problematic on several grounds, the most important being the switches of lateralization mentioned earlier. See Mohr 1976.) Neuroscience seems to have also cast doubt on some forms of Chomskian language acquisition in another respect, as well. While some aspects of the Chomskian program[26] are well established in psychology of language (Medin and Ross 1997; Pinker 1994; O'Grady and Dobrovolsky 1996), the hypothesis that humans are born with an innate language acquisition "device" seems implausible in the light of studies showing how the neurons a human child is born with are too uncommitted to have such a device "already built". New hypotheses in the epistemology of language acquisition have to use these neurological findings as constraints - so far very little work has been done in this area. The lesson for the philosopher here is not only in epistemology proper, but also how epistemology relates to a philosophy of language, and perhaps also to a "metaphysics of language"[27] (for example, McGilvray 1997 requires updating in light of this).

Now that we have seen how neuroscience gives epistemology new directions, let us look at the other way. This subsection mirrors the last as the last looked at neuroscience's influence on epistemology and this looks at the influence of epistemology on neuroscience. I shall examine this in two respects. One respect concerns the issue of reduction, which has been dealt with briefly earlier. The second respect concerns an issue at the border between epistemology and methodology. This concerns the question of **ecological validity**. The minimalist sketch I am presenting should keep both these as part of its lessons.

Reductionism is a central epistemological issue, but one that is a bit ambiguous. A relatively noncontroversial form of reductionism in neuroscience would be research into finding neural mechanisms of various cognitive and subcognitive processes.

Some philosophers (for example: Davidson 1980) have denied that there are any lawful correlations[28] between the psychological and the neurophysiological. These philosophers pose no threat to this weak form of reduction mentioned above, as their thesis is question begging against the

[26] Or some version of it, anyway. Chomsky's viewpoint has changed over the years.

[27] This branch of metaphysics concerns itself with what general features of the world (including those of humans and other potential language users) are necessary in order for language to be possible. The term "metaphysics of language" was coined by me in the light of the referenced work.

[28] Strictly speaking, Davidson's claim concerns the nonexistence of psychophysical laws. The two theses are equivalent for the purposes of the present paper.

discipline of cognitive neuroscience itself. It hence poses no further interest to my minimalism. What is important for the minimalism that should be examined is whether any strong reductionism in neuroscience and hence in epistemology is warranted.

A medium level version of reductionism is as far as I feel is currently warranted. Brain studies have localized parts of the mechanism for the Stroop effect, for instance (Sicklas 1998). This is an example of a psychological effect to be explained in terms of an underlying neurophysiology.

What is not warranted by the current findings in neuroscience is the **dispensability** of conventional psychological classification. The Stroop effect here is paradigmatic. To understand what is going on and the interest in it, both the psychological (or “phenomenological”) level and the neurological level which shows a mechanism for its effect are necessary. The neurological level alone here (as would be warranted by a strong case of reductionism) would not be interesting - it would likely be extremely difficult to tell that this organization of neuronal assemblies had this outcome in behaviour.

On the other hand, there may very well be cases where more reduction is useful, where the psychological level is misleading or worse. (This also suggests that a philosopher of neuroscience should be willing to encourage refinements at the level of conventional cognitive psychology.) One of these areas is vision. Philosophers have had a lot to say about vision and related issues in the context of the philosophy of perception. This is our next section, which we shall see after a discussion of ecological validity. But first, what is the lesson for the minimalist from this discussion of reductionism? There is no one single lesson, except for that one should always keep an eye out for how much reduction is tenable in each case, and not blindly assume that neuroscientific research supports a uniform amount of reductionism across the board.

Our final topic in epistemology and neuroscience concerns the issue of ecological validity. I propose to discuss it here as it concerns a general problem of knowledge at the root.

A scientific experiment is said to be ‘ecologically valid’ to whatever degree if it uses the subjects of the experiment in a situation that ‘mimics’ or is in some sense ‘like’ a ‘situation outside the lab’. The concern is over wondering how relevant lab studies are to real world contexts. This is especially important in psychology, as the situations created by experimenters are often very unrealistic, focused situations. While there are many facets of and solutions to this problem, here I shall focus on the ones that pertain to neuroscience.

It is very difficult at the present time to have maximal ecological validity in a neuroscience experiment these days. One cannot have the subjects in a brain imaging study walk around; they must stay in the MRI apparatus and so forth. Curiously enough, this limitation actually has some strengths as well as the weakness of limiting the experimenters to investigate certain kinds of activity. (It would be very difficult to determine with this method what brain centers are involved in playing certain sports, for instance.) It does have the advantage that it prevents irrelevant (to the current experiment) brain activity. For instance, if one is studying the effects of visual stimuli on various brain centers, one does not want the brain centers responsible for walking to be activated at the same time.

Being forced to use machines which limit ecological validity in neuroscientific investigations is thus a two edged swords. Epistemologists should take note that ecological validity concerns are simply a specific case of more general issues in the methodology of science. Laboratory manipulations are first approximations to “situations in the real world”, presumably. It would seem to beg the question against the scientific world view to take the position that what is neurologically or psychologically produced in subjects in laboratory experiments has <u>nothing</u> to do with the real world without in fact doing neuroscience in both places and showing this. This possibility does not seem likely, but one should not assert that it is necessarily not the case on a priori grounds. The principle of lawfulness does not necessarily help here, as neural assemblies in use in the brain may be very different at different times. The postulate of lawfulness would only rule out no regularities in what is used at a given time. The epistemologist concerned over ecological validity might be concerned that the situation of being evaluated in a laboratory may cause activation of different neuronal assemblies. The lesson for the minimalist, therefore, is not to rule out this possibility out of hand, and to encourage scientists to develop less intrusive ways of studying brain functioning so that the nervous system may be studied in “real world” situations as much as possible.

Section 1d - Philosophy of Perception

As remarked in the previous section, epistemology and the philosophy of perception are strongly related. Here I will discuss specific issues concerning several topics in the philosophy of perception and how they could be informed by a minimalism in the philosophy of neuroscience. I shall discuss briefly the issue of qualia (including colours), the issue of "privileged access" and finally the issue of mental representation. (Philosophy of perception stands somewhat at the border of epistemology and philosophy of mind. Therefore, if some of these issues seem to be out of place, I suggest that it is simply a matter of personal preferences when it comes to differences of classification.)

Qualia are the first topic that neuroscience can be brought to bear on in the philosophy of perception. First, let me make it clear that I am talking about qualia in the one of the stronger senses[29]. These are said to be 'what is left out' when all the representational aspects of a mental state are removed. Supposedly, these 'what is left out' are things like feels, visual fields with their apparent properties like colours, and so on. I move that neuroscience is able to cut the Gordian knot about qualia once and for all. Before this is done, however, an elucidation of what is meant by representational is required. I believe that part of the debate over qualia is precisely over the meaning of this word. Once the meaning is clarified, perceptual neuroscience takes care of the rest. The lesson for the philosopher is that her job is only part of the picture here.

So, what is a neural representation? I move that neuroscience has told us that structures like sentences are not likely represented in sentence-like fashion in the brain. Unlike a conventional computing device which may be said to have "The cat is on the mat" distributed sequentially[30] across some memory locations under certain circumstances, a brain has a more abstract representation that is not sentence like at all (see, for example, Feldman 1989; Churchland and Sejnowski 1989). It seems plausible then, that since we intuitively do not have any idea of what exactly is represented in

[29] The weaker sense of qualia, which is just "secondary property" is also contentious, but not relevant to our present purpose. Neuroscience is relevant here as well, but philosophical lessons are pretty unimportant here, since most philosophers take some sense of secondary properties seriously. The issue being discussed in the present work is somewhat similar to a discussion of whether colours are secondary properties.

[30] Here sequentially should be understood as logically sequentially. There is no reason why a conventional computing device's programs need store data items in physically adjacent cells.

our brain, simply asserting that certain ‘feels’ are strictly qualitative and aren’t representational in content needs support which is not forthcoming. The attempts of some philosophers to prove their theses (either for or against this thesis) without reference to clear neurophysiology of representation are hence misguided. (There is a possible source of misunderstanding here that I shall deal with in the second large section of this paper, below.)

One clear example of neuroscience showing the philosopher of perception that her prescientific concepts of perception are inadequate concerns the notion of a *visual field* which is ubiquitous in the philosophy of perception literature. (See, for example, Philosophical Issues 1995) However, this notion relied on is intuitive - we appear to have a “space” which sort of surrounds our head in which our visual experiences are presented. This field is said to be uniform, possessing a Euclidean metric, containing colours throughout, and so on - after all that’s what it APPEARS to be. Nevertheless, neuroscience (together with psychophysics) has shown otherwise, as I will now demonstrate. A pre-neuroscientific[31] explanation for the following demonstration would be completely ad-hoc. This demonstration does not require any fancy equipment or even a partner to take notes and is so ideal for encouraging certain philosophers to overcome their misconceptions. (I adapt this from Dennett 1991 with a few pointers from Sekuler and Blake 1994.)

The demonstration[32] proceeds as follows. Take a deck of playing cards, and shuffle well. Take one card from the top of the deck and turn it up away from you at arms length. Without exposing it to yourself, rotate one’s arm containing the card so that it is 90 degrees or so from the center of the eyes. Turn the card inwards so that its face is in the direction of one’s head, while fixating ahead. Ask yourself whether you even “think” there is a card there. Many people, including the present author, despite their knowledge of the peripheral vision system, were shocked to find that part of their brain was “saying”: “I don’t know what is out there. It isn’t a card. I can’t tell that.” Observe carefully one’s inner reactions for “views” like that, and finally, slowly, bring the card towards the center. Notice that it takes a while even before the card takes on a **shape** (that is, many people

[31] Note that in essence the nervous system extends to and includes the eyes. This fact about biology is also overlooked by some philosophers (Gold 1999).

[32] I encourage the reader to take time to do this, as it is really quite striking. Simply saying “oh, yeah, I know what happens in peripheral vision” is not enough.

report seeing that “there is something there, I know not what” at some point early on. Only later does the card appear to take on shape, followed in turn by colour and only quite near the center does it appear to take on an identity (as, e.g.: the four of diamonds.) This demonstration is striking even if one is used to the effect from previous demonstrations like it. The present author has done it many times and has been surprised every time.

What, then, is the lesson for philosophers from this demonstration? To borrow the title of a paper (See P. S. Churchland, Ramachandran, and Sejnowski 1994) on this subject, it is part of a “critique of pure vision”. In other words, most philosophers (see particularly Philosophical Issues 1995) have tied themselves in knots in debates over what is exactly the visual field contains. This in turn has fueled debates about representation, as we have seen above. The general lesson, for philosophers, is not just that our intuitions about such matters are grossly mistaken most of the time, but that we can in fact learn more about our own perceptual states than just by simply introspection, even if it is at a low level of sophistication. It is also important to draw Dennett’s (1991) lesson, between “really” and “really seeming”. Nobody is going to deny that it **seems** like we have a visual field as I described it above. However, the mistake is thinking “it”[33] is <u>really</u> like that. The fundamental lesson of neuroscience to philosophy of perception is precisely that there is this very large gulf between the way things are really with us and the way things seem to be with us. This is even more critical when it is noted that other, non-visual systems of the brain have large interactive connections with the visual system. We do not generally notice that our visual system has connections to the motor cortex, to the thalamus, to auditory centers[34], and so on. The functions of some of these paravisual systems are still unknown, particularly those that connect to subcortical structures. See P. S. Churchland, Ramachandran, and Sejnowski’s (1994) paper suggest that one should adopt a “interactive” vision approach. Previous neuroscientific, psychological, artificial intelligence-oriented and philosophical views on vision were much too “static” in character. The aforementioned paper (correctly) emphasizes the importance of the non-visual states of the organism in the context of what is perceived. A philosopher should be very wary about drawing conclusions in philosophy of perception based on what

[33] Part of what this another results along the same lines demonstrate is that it is somewhat misleading to talk of <u>the</u> visual field as if it were a “unified object”.

[34] This particular linkage was actually discovered at the cognitive/behavioural level first and later found on neuroanatomical investigation.

appears to him to be the case on introspective grounds.

This "veil of ignorance" extends to how some philosophers and neuroscientists have interpreted psychological phenomena such as blindsight (See Weiskrantz 1989). Dennett's (1991) book points out the danger of taking people at their word, especially when they have brain damage. One need not postulate other undiscovered visual pathways to explain blindsight. If one had a visual deficit and was having it mapped out, and the experimenter asked "where did you see it this time", this is producing a cuing effect. None of the blindsight subjects ever report anything in the absence of these cues. None of them volunteers anything like "Pardon me, did you happen to shine a light in my scotoma?" Again, the "anticartesian" lesson is clear - we simply do not have infallible access to the contents of our brains. Philosophers of perception should note that we are now at the stage to deal with qualia.

As noted above, there are at least two ways that this term is used in philosophical literature. The sense of qualia as secondary properties is relatively innocuous and will be ignored in the present work. However, the other sense has caused much philosophical ink to be spilled and must be dealt with in the light of our "neurophilosophic" look at representation. Could there possibly be anything above and beyond representation in the sense I have sketched? Apparently not, as I have discussed several ways in which representation is done by the brain - piecemeal, often in several ways at once (recall my report of "part of me" insisting that it did not know that it was a playing card in my peripheral vision) and so on.

Qualia are said that which is over and above representational content. But what brain subsystem is not representational? Not all systems need represent external states of affairs; many neuronal assemblies no doubt "monitor" things within us. However, there just does not seem to be any place for a strict "feeling" being produced. I am not denying that things appear to have certain internal feels or appear to have certain colours. I am denying that these are anything but "how the representation seems to us" - but this does not occur (*pace* Penfield (1975)) at any place in particular in the brain. Representation occurs all over. There is no need to posit a further "place where it all comes together".

It **seems** that a banana has a yellow "covering" we call a colour. It **seems** like an indication from the stomach to eat something is a "stomach ache like <u>this</u>". But there's no systems in the brain that do anything but represent - none “present” to an audience (the “self”). Hence there are no qualia in the contentious sense.

The second and final issue in the philosophy of perception I shall deal with concerns the relationship between philosophy of perception, neuroscience, and other sciences (particularly psychophysics). The study of perception does have a long philosophical tradition (Democritus wrote about it, among other ancient thinkers.), but that is insufficient for it to continue in the light of scientific studies of perception. What, then, is the role for a philosophy of perception in light of neuroscience?

I suggest that it be limited to three areas, which are not meant to be mutually exclusive. First, philosophers of perception can work as specialized philosophers of science - ones who understand the various branches of perceptual psychology, neuroscience, etc. and are able to perform tasks of clarification, elucidation etc. that philosophers of science generally do, but within this domain. Second, they can bring findings from the sciences of perception into philosophy, particularly into epistemology and philosophy of mind, to avoid dogmatism. (This is what I have tried to do in the discussion of representation and qualia, above.) I move that general critiques of perceptual science without an appropriate general philosophy of science backdrop are ungrounded and futile. (Witness the debate over qualia.) The third task for philosophers, which generally happens in spite of them, is the encouragement of interdiscipline building and particularly interdisciplinary theory building. This is somewhat subsumed in the first case, but can be done with a logic background as well. (Of course, logicians should carefully note “real world” reference classes and so forth that they are perhaps not used to dealing with.) These logicians would ferret out useful ways to represent formally certain neuroscientific results and hypotheses. In this sense they would work as mathematical neuroscientists with a philosophical ‘spin’ on things. (This latter bit is important, as presumably it would give them insight into the metaphysical and epistemological issues in question.)

Neuroscience informs the three kinds of philosopher of perception in the various ways I have outlined above. It also restricts her traditional job description somewhat extensively. It is important to realize that this is bound to happen with the advancement of science, and is not something that should be too strongly resisted. It should not cause too much of a problem, as there are still many very interesting areas in this field that do require the attention of philosophers.

Section 2 - Responses to Possible Critics

I have dealt with some forms of criticism already in section 1. However, contemporary dualists or others that are skeptical of neuroscience telling us about mentality and so forth have been given a bit of a short shrift. I shall therefore in this section discuss objections to the minimalist program from the eliminativist camp, from the emergentist camp, from the contemporary dualists and related positions, from those who think another, completely new discipline is required in this area and neuroscience is missing something unless it includes or makes use of it (Penrose 1989,1994, 1997) and a brief mention of some possible more directly scientific criticism. Since my paper has been primarily aimed at philosophers, I do not feel that I need dwell on the last point too much. This is because this paper has not made any particularly far reaching scientific (or metascientific) claims[35].

Contemporary dualists (of which there are some in both the philosopher camp and the scientist camp) can be classified into two main categories. Those that call themselves dualists[36] or fully admit to their unorthodox position (such as Swinburne, Eccles, Penfield[37], etc.) are those in the first category. There are also positions which are so called tacitly dualist. Fodor’s (1980) ‘concession’ that there might be minds made of (or instantiated in) non-material stuff is one such example. I shall ignore the tacit dualists and focus on those that hold explicitly dualist views, as determining who exactly holds the other views is difficult to determine. Some of their objections will be dealt with in due course under other categories of objections anyway, so this should not matter terribly.

[35] Note that while this paper is a paper in the philosophy of science, the minimalist viewpoint is a call for caution in the rush to adopt neuroscientific findings into philosophical discourse. Hence the view of the present author is more ‘conservative’ than many in this area. (Particularly as compared to the Churchlands.)

[36] While materialism is on the upswing amongst philosophers, dualism is still influential and still has partisans. As the list of dualists indicates, some prominent scientists have also been such, as well. Hence I feel that dealing with some of the more cogent or important (because of the potential influence of the holder of the view) dualist arguments against neuroscience is useful. It also clarifies some of the positions sketched in the minimalist program of the first section of the paper. Opposition between diametrically opposing views often yields insight, so I will pursue these. I am still of the opinion that any sort of dualism is completely incompatible with science, though.

[37] Penfield denies that he is a dualist, but adopts which is dualist by any accepted definition of the term. Hence he is guilty of the “fallacy of redefinition.”

I will deal with the objection to neuroscientific explanations of consciousness of Penfield (1975). (I personally find Penfield to have been the most baffling of all the dualists, given all the work he did in understanding brain function.) Penfield writes that he cannot find the location in the brain for “the mind”, and suggests to his readers that is because it is nowhere in the brain to be found. This of course is a horrible argument, one that would not get even a grade of “D” in an introductory philosophy course. We should hence not take Penfield’s philosophical speculations under any further consideration[38], but before I leave Penfield’s contribution to philosophy of mind through neuroscience behind, I would like to deal with an possible objection, which follows.

It may be correctly rejoined that Penfield was not a philosopher and should not be judged by philosophical standards. Two equally important answers to this rejoinder may be given, as it seems to be a non sequitur. One is that *The Nature of the Mind* (Penfield 1975) explicitly states that it is dealing with to philosophical issues (see the preface). The second is that the mind-body problem generally is a philosophical issue in part (see (e.g.) Bunge 1980; Douglas 1998b; Dennett 1991; P. M. Churchland 1995 for long lists of reasons why the mind-body problem is partially philosophical and partially scientific) and so claiming that a work on it is purely scientific and should not be judged by philosophical standards is wrongheaded. By the same token, a strictly philosophical work on the mind-body problem is equally wrongheaded.

Another objection to neuroscience comes in the context of attacks on techniques. It has been pointed out that while a neuroscientific investigation can produce evidence that some brain processes are “mental”, no conceivable experiment could show that mental processes are all such. (Even in the case of restricting the universe of discourse to humans.) This is, strictly speaking correct. However, it does not make a wit of difference to scientific explanation. This is basically a rephrasing (in the context of neuroscience) of the problem of induction, which as many authors have pointed out, ceases to be a problem if one goes beyond the Baconian idea that a scientific theory is just a data summary. (Of course, this

[38] Of course, the argument is a bit more complicated than I have presented it - however, it is in essence an argument from ignorance. (There is also a bit of an appeal to pity, as he recounts the story of a Soviet physicist he was asked to treat making a somewhat remarkable recovery that Penfield did not understand. He (Penfield) attributes this recovery to the action of the mind. The pity comes in when he asks the reader to consider the plight of his patient and “how it was possible” and so on.

seventeenth century idea is still somewhat popular amongst philosophers, so perhaps another lesson to learn from neuroscience is that this thesis is false.) In essence, the objection begs the question against scientific explanation, which is precisely what is at issue.

Another objection we must deal with from the dualists concerns qualia. We have seen there do not appear to be any qualia in the contentious sense by investigating the brain. However, it may be rejoined that some (particularly Nagel 1974) are arguing against materialism with their views on qualia, and hence I have begged the question against them by dismissing the views by reporting on brain investigation. We have seen a version of this argument in section 1b (ontology). Now that we have seen some remarks on philosophy of perception (see section 1d), let us look at this dualist argument in a bit more detail - one particularly influential one follows.

Qualia are said not to be brain states because they are “out there”, and that (e.g.) “reddish-orange afterimages” are reddish orange and a brain process is acoloured (is not the kind of “thing” to have a colour). So the qualia of the afterimage is not a brain process. The best answer to this is the **seeming** card we learned about from Dennett. It seems as if my afterimage is reddish-orange, but there isn’t really anything there, so nothing is really reddish-orange. The neuroscience even gives us the beginning of an explanation for the “existence” of afterimages (or rather, why they appear to us), to boot. (See Sekuler and Blake 1994 for some of the basics.)

Some philosophers have also thought it might be possible for all the brain states of a person to be the same as another, but the “how it seems to them” not be the same between the two. Often times, this objection is cached out in terms of the difference between a functional (identified here with a materialistic (or “physicalist”, as it is often put) account and a phenomenological account. This objection is too question begging to take seriously, despite its popularity, particularly amongst those who indulge in worrying about “possible worlds” and so on[39]. I will ignore it and leave the dualists behind and move on.

Let us move on then to a **slightly** more sophisticated attack on neuroscience, that of Roger Penrose. Penrose thinks that a new revolution in physics will be necessary to understand the brain’s functioning (Penrose 1989, 1994, 1997). His (apparently) strange (but not dualistic) thesis has

[39] A brief exploration of this claim of mine is found in Douglas 1998c; it would go too far afield to present it here.

attracted some philosophical commentary and some responses from various scientists of the brain and behaviour. I will here examine his claim about the "new physics of neuroscience[40]" very briefly, and then explain several ways to deal with this objection.

Penrose attempts to show in his works on consciousness that understanding it will require a new revolution in physics. He claims that consciousness requires an instance of a **noncomputable process**. This is taken in the strong sense[41] as a process that cannot even be simulated (to any desired degree of accuracy) with "computable functions" (here also taken in the mathematical sense).

He asserts that the collapse of the wave function in quantum mechanics is precisely such a process, and then proceeds to try and find something in the human nervous system that will collapse wave functions. He suggests that microtubles in cells may work in the "special way" he is proposing is necessary. The minimalist program above (section 1) would be falsified if Penrose's speculations were to be born out, as one of the principles of the minimalist program is the metaphysical principle of level adjacency. (I.e. that one should not skip ontological levels in scientific explanation.)

Numerous objections can be made to this scenario. Since the present work is a paper on the philosophy of neuroscience, I shall only produce arguments bearing directly on the neuroscientific speculation. Penrose, as we have seen, suggests a quantum effect in microtubles producing consciousness (he is actually a bit more specific than this, but this is the

[40] I am hereby dividing his claims about non-algorithmicity in human thought, Gödel's Theorem, the Halting Problem, and so on from his specifically neuroscientific claims. For discussion of the former issues, see Douglas 1998a for a summary of responses and criticisms.

[41] A weakly noncomputable process can be simulated with any desired degree of accuracy with computable processes. For instance, calculation of the value of a transcendental function is weakly noncomputable. Very often computer scientists do not use the terms from the mathematical theory of computation (there is also a "physical" or "computer science-oriented" theory of computation, which is a bit different) so there is some confusion in this area amongst those who argue against or with Penrose on this issue. It is not known at this time whether there are any strongly non-computable processes. The Church-Turing thesis can be read as the assertion that there are none. It remains to be seen whether all connectionist models of the brain are Turing equivalent (i.e. weakly computable), and further to see if the brain's functioning is weakly computable. Should it turn out that the brain's functioning is not weakly computable, connectionism, and indeed, any modeling, description, or theory building involving conventional mathematics is misguided. Penrose's claim is thus extremely strong.

essence of the position). This provokes the question: why do quantum effects (if there are any) of microtubles in brain cells have this effect, and not the microtubles of livers or other parts of the body. (As, after all, microtubles are far from being unique to neurons.) Also, other animals have microtubles in their cells (in fact, so do plants) - are they conscious[42] too? The Churchlands (see Churchland & Churchland 1998) have also pointed out that the quantum mechanical effect Penrose wants can be easily washed out by the presence of certain ions which would occur relatively often in the microtubles of certain cells. They also point out that a certain medicinal chemical used once interfered with microtubles but had no effect on the subject’s consciousness. It is, therefore, pretty much safe to say that Penrose’s current neurological speculations are completely at odds with what is actually known about biochemistry and biophysics.

Enough of the objections that are far removed from neuroscience. I shall now deal with objections from a possible eliminativist. Since the most notorious of the eliminativists are the Churchlands, and in the interest of space and time of this paper, I shall make remarks on their possible objections exclusively.

The first objection to deal with concerns the nature of what I would propose has successfully “eliminated” by the progress in neuroscience. Alternatively, this can be reworded as asking whether I recognize any full eliminations - the example I gave above of memory was only partial. I move that the pop-psychological category of the sense of touch has been eliminated, except in so far as one can speak of “sense apprehended through the skin”. This cannot be a adequate definition of touch, as we can have sensations (in the prepsychological sense, anyway) of touch apprehended through the eyeball, the tongue, etc. I would move that touch has been eliminated in favour of pressure, heat, etc. senses.

The eliminativist may then ask me how the minimalist program has been rather vague on when it occurs and when it doesn’t; and further that perhaps it still will occur in notions that I have claimed have been shown to be legitimate neurophysiological processes. Let us take the example of “emotion” to draw out this controversy. A emergentist will claim that emotions are processes involving in part (e.g.) the limbic system. This is true in so far as it goes. But what I claim the minimalist should adopt is

[42] This fact is even more damaging to Penrose if one recalls that part of the motivation behind his account is the human ability to do mathematics. Surely cats, sponges, and perhaps even roses are not mathematicians.

precisely this finding. I do not claim to know that all of the emotions we prescientifically recognize will be ‘natural kinds’ locatable in specific brain functionings. But even if they all (taken individually) are not so locatable, that does not mean necessarily that the notion of emotion itself is thereby eliminated as it may still refer to higher order patterns.

I shall now finish this philosophical section of responses to objections by dealing with objections from the emergent materialists, notably Bunge. My own minimalist account can be characterized as one version of emergent materialism, but of a broader sort than that of Bunge’s. As we have noted above, I allow for the possibility of minds being the functions of other things than plastic nervous systems. I will deal with 3 objections to this account and show how they do not follow from what is known about in neuroscience. (If it is claimed that they are not supposed to, the point is granted, however, those who often endorse such claims provide support for these objections on neuroscientific grounds. As we shall see, most of the reasons given are acceptable, but do not have the desired consequences.)

The first concerns the minimalism’s account of what can have properties of mentation. Neuroscience is informing us of how brains composed of neurons have emergent mental properties. So, it appears then that we know what mental properties are - that they are brain processes. I agree with this claim. Mental processes are indeed brain processes, in our case. From that it cannot be concluded that nothing else can produce mentation via emergence mental properties. It seems that from what we know from neural modeling that the substratum for the neurons is actually irrelevant, and that only their inputs, internal states, and outputs matter. It remains an open question to see if these models can be constructed at large enough scales to have more capability for modeling higher order processes. But the question remains - at some point, do these systems cease being simply models and become systems actually capable of learning? There seems to be no principled way at this stage to deny that such systems learn[43]. It would be very strange indeed if it turned out that “wetware” was required to produce mental functioning, as that would (at least with present knowledge of biochemistry and the like) be a “spooky” sort of emergence. Why is this so? Well, it appears from models like those presented in Nadel *et. al.* 1989 and Koch & Segev 1998 that some aspects of the chemistry and certain specific features of the biology involved are irrelevant functionally speaking. Note also that putting things this way does not commit one to any sort of strong

[43] The person who wants to deny this has to give some reason why the models do not learn that does not merely stipulate the possibility away.

functionalism[44]. So then it appears that it is an open question whether systems of non-biological “neurons” are possible.

The second objection from the emergentist to deal with concerns my cautiousness. The emergentist quite rightly claims that many mental functions have had their neural systems discovered and explored. So shouldn’t the minimalist doctrine I am espousing be more than minimalist in these areas? Shouldn’t we adopt these (provisional) findings, until new notice? This is in many respects the “mirror image” of the objection by the eliminativist discussed previously. The answer is essentially the same - each case is to be handled by itself, on its own merits. It should also be noted that it isn’t clear how one is to interpret the emergence here. Yes, it is true that certain properties of systems of neurons are emergent in the (relatively) non-controversial sense (Bunge 1977). However, one has to be very careful (as noted above) of the redefinition fallacy. Because “pop-psychological” terminology is vague, any proposed emergent property of neuronal systems to explain the psychological category it refers to is bound to sew some confusion due to this. An example would be emotion. Limbic functions and their properties are certainly involved in the “production” of emotion; however are they identical with it? How does one tell? The emergentist must specify before hand what her reasons and standards of evidence are to be, otherwise, she will battle fruitlessly with the eliminativist and the traditional cognitive psychologist to no end. The lesson for the philosopher here is that there is a lot to consider before adopting a thesis of emergentism in any particular. The present author is of the opinion that the thesis in general is a good one.

The third and final emergentist objection to the minimalist program concerns the account of qualia in the first sense. I have said that the secondary property meaning of qualia is not controversial. An emergentist may rejoin that these secondary properties are some of the properties that he is talking about when it comes to his positions on neuroscience. I think this is correct, that secondary properties are emergent properties involving brains. However, I do not think that the emergentist should be too hasty here, which is why I am not building these into the account of minimalism. The minimalism I have been sketching is the amount of <u>neuroscience</u> that is

[44] Strong functionalism would be the thesis that mentating systems can be made of any material whatever. I think that stone cannot support such a system, for instance, so I am not a strong functionalist. (Of course, like any real world dichotomy, the distinction between strong and weak functionalism is a distinction used to delimit “poles” of a whole set of possible positions.)

relevant to philosophy. To build in an account of secondary properties would involve building in a general ontology, which while useful, is not directly related to the present project.

Another objection that must be dealt with concerns the minimal computationalism discussed above in section 1d above. The objection would run something like this: it is true that the models you propose are computational in character; however, how does this support the thesis that the nervous system itself is computational in the relevant respect? Further, how does one deal with issues below the neuronal level on this account? Surely the cellular and extra-cellular neurochemistry is important somehow. This family of objections can be answered several ways by an adopter of the minimalist account. First, it can be pointed out that many very recent attempts to model computationally various aspects and systems of the nervous system have attempted to build in various cellular chemistry level details (Koch & Segev 1998). Second, it isn't immediately clear how this level is relevant. This is related to the previous discussion on functionalism. This question is as yet unanswered. The first answer I have given to these objections should remind the reader that work is proceeding in this area regardless. Thirdly, and most related to the "really" computational issue is the following. There has been enough work in recent years in the <u>computational</u> aspects of neuroscience (as understood in the constraints way discussed above) to suggest the ball is now in the court of those who oppose this method of constraint to research. What these partisans must show is that (a) this approach is misleading or worse, and (b) their objections to the approach overcome the problem of stipulation. In other words, they must show that it is more than a matter of stipulation what counts as a computational process in order to rule out the possibility of understanding nervous systems this way. (I, based on Dennett's **intentional stance**, call this matter of stipulation the **computational stance**. The minimalist suggests that some amount of the computational stance is necessary to understand the functioning of the nervous system.)

The objections dealt with in this section taken together have one common moral, however, which is important to point out in order that the future of the philosophy of neuroscience is assured. This moral can be stated in one word: "balance". The balance should be between excessive arm chair scientizing and excessive arm chair philosophizing.

<u>Section 3 - Conclusion & Thanks</u>

In the present paper we have seen four large branches of philosophy and their influence on and by neuroscience. I have warded off several objections from without and attempted to set a few of the controversies within the philosophy straight. Finally, we have seen a beginning of how future research in philosophy of neuroscience should go, as a balance between two extremes of emergentism and eliminativism, with a tiny amount of a weak functionalism and computationalism thrown in.

Thanks go out to Ian Gold of philosophy and Professor Petrides of McGill’s psychology department who got me thinking more about how neuroscience interacts with philosophy. This paper could not have been written without the influence of their stimulating courses.

Appendix A - The Stroop Effect

This is an interesting psychological fact about how conflicting “messages” can cause interference in performance. Subjects are asked to name the colour of each word aloud as fast as possible in a sequence of words like the following:

BLACK
GREY
GREY
BLACK
BLACK
GREY
BLACK
GREY
BLACK
GREY
GREY

The reader should confirm for himself that he does indeed have to slow down to avoid making mistakes or, should he still try to name at full speed, that he indeed makes many mistakes.

References

Beardsley, T. 1997. The machinery of thought. Reprinted in Petrides 1998.
Bunge, M. 1973. *Philosophy of Physics.* Dortrecht: Reidel.
Bunge, M. 1974. *Sense and Reference*. Volume 1 of Treatise on Basic Philosophy. Dortrecht: Reidel.
Bunge, M. 1977. *The Furniture of the World.* Volume 3 of Treatise on Basic

Philosophy. Dortrecht: Reidel.
Bunge, M. 1980. *The Mind-Body Problem*. Oxford: Pergamon Press.
Bunge, M. 1983. *Exploring the World.* Volume 5 of Treatise on Basic Philosophy.
Dortrecht: Reidel.
Bunge, M. & Ardila, R. 1987. *Philosophy of Psychology*. New York: Springer.
Churchland, P. M. 1995. *The Engine of Reason, The Seat of the Soul.* Cambridge: MIT Press.
Churchland, P. M. & Churchland, P. S. 1998. *On the Contrary*. Cambridge: MIT Press.
Churchland, P. M. & Sejnowski, T. 1989. Neural Representation and Neural Computation. In Nadel *et. al*. 1989.
Churchland, P. S., Ramachandran V, & Sejnowski, T. 1994. Critique of Pure Vision. In Koch and Davis 1994.
Churchland, P. S. & Sejnowski, T. 1996 [1992]. *The Computational Brain*. Cambridge: MIT Press.
Davidson, D. 1980. *Essays on Actions and Events*. New York: Oxford University Press.
Dennett, D. 1991. *Consciousness Explained.* Boston: Little, Brown & Co.
Dennett, D. 1998. *Brainchildren: Essays on Designing Minds*. Cambridge: MIT Press.
Douglas, K. 1998a. Some Responses to Critics of AI. Unpublished paper for McGill University course 107-341B, Winter 1998 semester.
Douglas, K. 1998b. Psychoanalysis and Pseudoscience. Unpublished paper for McGill University course 107-590A, Fall 1998 semester.
Douglas, K. 1998c. Academic Games and Kripke's Antimaterialism. Unpublished paper for McGill University course 107-306A, Fall 1998 semester.
Feldman, J. 1989. Neural Representation of Conceptual Knowledge. In Nadel *et. al*. 1989.
Fodor, J. 1980. The Mind-body Problem. Scientific American[45].
Gold, I. 1999. Unpublished lecture for McGill Course 107-419B, Winter 1999 Semester.
Gold, I. & Stoljar, D. 1999. A Neuron Doctrine in the Philosophy of Neuroscience. Forthcoming in *Trends in Neurosciences*.
Hamlyn, D. 1984. *Metaphysics*. New York: Cambridge University Press.
Hebb, D. 1949. *Organization of Behavior.* New York: Wiley.
Hebb, D. 1980. *Essay on Mind*. Hillsdale: L. Erlbaum Associates.
Hofstadter, D. 1985. *Metamagical Themas: Questing for the essence of mind*

[45] I was made aware of this article by Professor D. Davies of the McGill University philosophy department. I have not been able to track down the original publication.

and
pattern. New York: Basic Books.
Jackson, F. 1982. Epiphenomenal Qualia. Philosophical Quarterly 37, no. 127.
Koch, C. & Davis, J. (eds.) 1994. *Large Scale Neuronal Theories of the Brain.* Cambridge: MIT Press.
Koch, C. & Segev, I. (eds.) 1998. *Methods in Neuronal Modeling: From Ions to Networks.* (2e.) Cambridge: MIT Press.
McGilvray, J. (ed.) 1997. *Metaphysics: Innate Ideas*. Course pack for McGill University Course 107-421A, Fall 1997 semester.
Medin, D. & Ross, B. 1997. *Cognitive Psychology* (3e). Orlando: Harcout Brice College Press.
Messer S. *et. al.* 1988. *Hermeneutics and Psychological Theory.* New Brunswick: Rutgers University Press.
Milner, B. & Petrides, M. 1984. Behavioural Effects of frontal-lobe lesions in man. Reprinted in Petrides 1998.
Mohr, J. 1976. Broca's Area and Broca's Aphasia. Reprinted in Petrides 1998.
Nadel, L., Cooper L., Culicover, P. & Harnish, R., eds. 1989 *Neural Connections, Mental Computation*. Cambridge: MIT Press.
Nagel, T. 1974. What is it Like to Be a Bat?. Philosophical Review 83, no. 4.
O'Grady, W. & Dobrovolsky, M. 1996. *Contemporary Linguistic Analysis: An Introduction* (3e). Toronto: Copp Clark.
Penfield, W. *et al.* 1975. *The Mystery of the Mind*. Princeton: Princeton University Press.
Penrose, R. 1989. *The Emperor's New Mind.* London: Vintage.
Penrose, R. 1994. *Shadows of the Mind.* London: Vintage.
Penrose, R. 1997. *The Large, the Small, and the Human Mind*. Cambridge: Cambridge University Press.
Petrides, M. (ed.) 1998. *Human Behaviour and the Brain*. Course pack for McGill University Course 204-311A, Fall 1998 semester.
Philosophical Issues 1995. *Philosophy of Perception*[46].
Pinker, S. 1994. *The Language Instinct.* New York: HarperCollins.
Pinker, S. 1997. *How the Mind Works.* New York: W. W. Norton & Co.
Ramsey, P., Stich S, Rumelhart, D. (eds.) 1991. *Philosophy and Connectionist Theory.* Hillsdale: L. Erlbaum Associates.
Rasmussen, T. & Milner, B. 1975. Clinical and surgical studies of the

[46] I was made aware of this collection of papers by Ian Gold, then of the McGill University philosophy department. I have not been able to track down the original publication.

cerebral

speech areas in man. Reprinted in Petrides 1998.

Searle, J. 1992. *The Rediscovery of the Mind*. Cambridge: MIT Press.

Sekuler, R. & Blake, R. 1994. *Perception* (3e). New York: McGraw-Hill.

Sicklas, V. 1998. Unpublished lecture for McGill University Course 204-311A, Fall 1998 Semester.

Squire, L. & Zola-Morgan, S. 1990. The medial temporal lobe memory system. Reprinted in Petrides 1998.

Taylor, C. 1988. Wittgenstein, Empiricism, and the Question of the ‘Inner’: Commentary on Kenneth Gergen. Printed in Messer *et. al.* 1988.

Weiskrantz, L. 1989. Blindsight. Reprinted in Petrides 1998.