THIRD WRITING ASSIGNMENT

Please email me your assignment in MS Word format by Monday, November 7 **at noon**. Assignments turned in after that are considered late, so plan accordingly! Late assignments are penalized 1/3 grade for each day late.

I am looking for clear, concise explanations — for most questions, a medium-sized paragraph or two should suffice. (For the "paragraph or two" rule, subquestions like a., b. count as different questions.) See the guidelines below.

1. Assume your reader is familiar with the Mary in the black and white room example, but not Jackson's argument, not what the example shows.
   a. Explain Jackson's argument that the Mary example shows that what it's like to see red involves something non-physical.
   b. Explain the Old Fact reply and the Property Dualism rebuttal.

2. Recall our class discussion of Nagel's "What is it like to be a bat?" Some people contended by the end of our discussion that our ability to imagine being a bat was, to some extent, beside the point. Even if we could, it wouldn't resolve the deep mystery that is Nagel ultimate concern.

   At the end of the article, Nagel declares:

   > At present we are completely unequipped to think about the subjective character of experience without relying on the imagination — without taking up the point of view of the experiential subject. (224)

   How does this statement bear on his contentions earlier in the paper that we don't know what it's like to be a bat (e.g., 220–21)? Does it provide a way to answer the challenge we raised in class?

3. It seems that we can all conceive of water without H_2O, Sam Clemens punching another guy, Mark Twain, in the face, or heat turning out to be something other than MMKE (mean molecular kinetic energy). It also seems that we can all conceive of minds without brains and vice versa (or, more specifically, mental state M without brain state B, and vice versa). Explain Kripke's views on conceivability and possibility.
   a. Explain Kripke's contention that scientific identities like water = H_2O are necessary.
   b. Explain why our ability to conceive a situation in which something other than H_2O — say element W — runs in the lakes and streams, falls from the sky, and is essential to life does not show that water = H_2O is not an a posteriori necessity

(that is, how does Kripke explain away the apparent conceivability that water ≠ H2O?);

c. Briefly describe how Kripke's view differs from Smart's reply to his Objection 2 (page 63).

d. Explain why the possibility of your zombie twin poses a problem for physicalism. (Your zombie twin is a creature that is an exact physical duplicate of you yet is not conscious — it has no sensations or feelings at all; there is nothing it is like to be this creature. It seems that we can all conceive of our zombie twins.)

e. Explain Kripke's reasons for thinking the strategy employed in b) will not work in to explain away the apparent conceivability of your zombie twin.

4. In later work, *The Mystery of Consciousness*, Searle remarks that, if anything, his original Chinese Room argument was *too concessive*. He writes,

> You could never discover computation processes in nature independently of human interpretation because any physical process you might find is computational only relative to some interpretation. (*Mystery of Consciousness*, 16)

I have attached a much longer excerpt from Searle's article at the end of this assignment. Explain Searle's point from the excerpt and critically object.

By "critically object" I mean that you must object to Searle in your critical discussion. Even if you think Searle is absolutely right, I want to hear your best attempt to criticize his points. (If Searle's right then this piece of paper is a computer. Pay me $200 for it, instead of wasting $1,500 on Dell or Apple!)

Guidelines

Please use a large, easy to read font (12 point); double spacing; standard margins; page numbers; correct spelling and grammar.

The material in this third assignment is more complex than in previous assignments, but with the exception of the last question, it's still all material we have explicitly covered in lecture. Your grade is going to be a function of the ***quality*** of your explanation: how ***clearly*** **have you explained the concepts, distinctions, views, arguments, or objections?**

- I want to see evidence that you understand the material. Clear, well-structured writing is excellent evidence of your mastery of the material. In class you sometimes know that you have a thought or question — you know what you want to say — but you can't quite put it into words. In your writing you should aim for **clarity**: aim for finding just the right words.
- The intended audience for your answers is neither me nor the other students in the class — you know we are familiar with the view and the vocabulary in which it is stated. Your aim is rather to make the view, distinction, or argument easily understandable to someone *completely unfamiliar* with the material. Pretend your reader is a fellow phil/mind student who missed all the lectures on Searle, Jackson,

and Kripke, and your assignment is all they have to catch them up. If you introduce a bit of new terminology you think your fellow student won't know, you should explain what it means.

  - ⇨ You may assume your reader is familiar with basic logical and philosophical vocabulary; e.g., you may assume your reader knows what it is for an argument to be *valid*, or *sound*. Examples of other basic vocabulary: concept, property, necessary & sufficient conditions, implication, evidence. (Of course these terms have a very specific meaning in philosophy, so while you don't need to define them in your writing, you should make sure you use them correctly. Check the Routledge Encyclopedia of Philosophy if you're unsure about a term — see course website for link.)

- By far the best way to show that you genuinely understand the material is to express the view, distinction, or argument **in your *own words***, using your own examples to illustrate them if necessary. If you just paraphrase the lectures or readings, that shows only that you have the fairly low-grade skill of paraphrase, and not that you genuinely understand the material. Do not use quotations, unless you think a crucial claim either is so dense or so confused that it has to be unpacked word-by-word.
- Try to avoid loose use of logical language ("therefore", "thus", "it follows", "prove", "refute", "false", "true"). If you mean to say that a point or a claim is *true*, do not say that it is *valid*. Only arguments can be valid. Do not use "thus" or "therefore" or "it follows" to make assertions or state opinions; these words should be reserved for stating the conclusion of a chain of reasoning.

You may find it helpful to consult Jim Pryor's paper writing guidelines (see the course website for link). Even if done a great deal of philosophical writing, you'll benefit from reading these guidelines.

Searle, J. 1997. *The Mystery of Consciousness*. New York: The New York Review of Books, pp. 14–17.

It now seems to me that the Chinese Room Argument, if anything, concedes too much to strong AI in that it concedes that the theory is at least false. I now think it is incoherent, and here is why. Ask yourself what fact about the machine I am now writing this on makes its operations syntactical or symbolic. As far as its physics is concerned it is just a very complex electronic circuit. The fact that makes these electrical pulses symbolic is the same sort of fact that makes the ink marks on the pages of the book into symbols: we have designed, programmed, printed, and manufactured these systems so we can treat and use these things as symbols. Syntax, in short, is not intrinsic to the physics of the system but is in the eye of the beholder. Except for the few cases of conscious agents actually going through a computation, adding 2+2 to get 4, for example, computation is not an intrinsic process in nature like digestion or photosynthesis, but exists only relative to some agent who gives a computation interpretation to the physics. The upshot is that computation is not intrinsic to nature but is relative to the observer or user.

This is an important point so I will repeat it. The natural sciences typically deal with those features of nature that are intrinsic or observer-independent in the sense that their existence does not depend on what anybody thinks. Examples of such features are mass, photosynthesis, electric charge, and mitosis. The social sciences often deal with features that are observer-dependent or observer-relative in the sense that their existence depends on how humans treat them, use them, or otherwise think of them. Examples of such features are money, property, and marriage. A bit of paper, for example, is only money relative to the fact that people think that it is money. The fact that this object consists of cellulose fibers is observer-independent; the fact that it is a twenty-dollar bill is observer-relative. As you read the sheet of paper in front of you, you see certain ink marks. The chemical composition of the ink marks is intrinsic, but the fact that they are English words, sentences, or other sorts of symbols is observer-relative. My present state of consciousness is intrinsic in this sense: I am conscious regardless of what anybody else thinks.

Now, how is it with computation? Is it observer-independent or observer-relative? Well, there are a certain limited number of cases where conscious human beings actually consciously computer, in the old-fashioned sense of the word that, for example, they compute the sum of 2+2 and get 4. Such cases are clearly observer-independent in the sense that no outside observer has to treat them or think of them as computing in order for them to be actually computing. But what about commercial computers? What about the machine in front of me, for example? What fact of physics and chemistry makes these electrical pulses into computational symbols? No fact whatever. The words "symbol," "syntax," and "computation" do not name intrinsic features of nature like "tectonic plate," "electron," or "consciousness." The electrical impulses are observer-independent; but the computational interpretation is relative to observers, users, programmers, etc. To say that the computational interpretation is observer-relative does not imply that it is arbitrary or capricious. A great deal of effort and money is spent to design and produce electric machinery that can carry out the desired kind of computational interpretation.

The consequence for our present discussion is that the question "Is the brain a digital computer?" lacks a clear sense. If it asks, "Is the brain intrinsically a digital computer?" the answer is trivially no, because apart from mental thought processes, nothing is intrinsically a digital computer; something is a computer only relative to the assignment of a computational interpretation. If it asks, "Can you assign a computational interpretation to the brain?" the answer is trivially yes, because you can assign a computational interpretation to anything. For example, the window in front of me is a very simple computer. Window open = 1, window closed = 0. That is, if we accept Turing's definition according to which anything to which you can assign a 0 and 1 is a computer, then the window is a simple and trivial computer. You could never discover computation processes in nature independently of human interpretation because any physical process you might find is computational only relative to some interpretation. This is an obvious point and I should have seen it long ago.

The upshot is that Strong AI, which prides itself on its "materialism" and on the view that the brain is a machine, is not nearly materialistic enough. The brain is indeed a machine, an organic machine; and its processes, such as neuron firings, are organic machine processes. But computation is not a machine process like neuron firing or internal combustion; rather, computation is an abstract mathematical process that exists only relative to conscious observers and interpreters. Observers such as ourselves have found ways to implement computation on silicon-based electrical machines, but that does not make computation into something electrical or chemical.

This is a different argument form the Chinese Room Argument, but it is deeper. The Chinese Room Argument showed that semantics is not intrinsic to syntax; this shows that syntax is not intrinsic to physics.