

Searle's Argument Against AI and Emergent Properties

By Paul Almond, 28 December 2008

Website: www.paul-almond.com
Email: info@paul-almond.com

© Copyright Paul Almond, 2008. All Rights Reserved.

Searle's Argument Against AI and Emergent Properties

By Paul Almond, 28 December 2008

Abstract

John Searle argues against the idea that an appropriately programmed computer would be conscious by saying that consciousness is an emergent property of a physical system, caused by a particular kind of physical process, and that you should regard consciousness as likely when the kinds of physical processes causing your own consciousness are occurring. According to this argument, there is no reason to think that a computer is likely to be conscious, irrespective of its behaviour, because that behaviour is not being caused by physical processes known to cause consciousness. This is a flawed position. While it is reasonable to regard consciousness as an emergent property of a physical system there is no profound sense in which it can be said that different people's brains work according to the same kinds of processes and an appropriately programmed computer and a human brain would work according to different processes. Any difference between these situations is just a matter of degree and any argument that we should presume other people conscious because their brains work in basically the same sort of way could also be used to justify presuming an appropriately programmed computer conscious.

Introduction

John Searle is a philosopher well-known for taking a position against strong artificial intelligence (strong AI) [1]. Strong AI, at least according to Searle's classical definition, is the idea that a system is conscious if it performs the correct computations. Searle makes some arguments that strong AI is wrong and even incoherent [1,2,3]. I will not be trying to refute all those arguments here. This article is about Searle's assertion that consciousness is an emergent property of a physical system and that this makes it unlikely that an appropriately programmed computer, just by performing the correct computations, would be conscious. "Emergent property" simply means some higher level property that a system has due to what is physically happening in it at a lower level.

An example of this, which I will use throughout this article, is wetness. Wetness arises due to the low level behaviour of molecules in a system. There is nothing supernatural about wetness: it is simply a high level property associated with low level processes and it disappears when we look closely enough at the molecules. Using a computer to simulate the behaviour of molecules in a fluid would not make real wetness. Our simulation would be just that – a simulation. We would not get real wetness because our simulation does not involve the actual processes needed to cause real wetness. To cause that you need the particular physical process of molecules moving around in a fluid. The physical processes underpinning the simulation are completely different: they are just

movement of electrons through semiconductor junctions. Electrons are not wet and movement of electrons does not cause wetness. Searle makes a similar argument about consciousness, saying that a simulation of a mind should not be expected to duplicate the consciousness of that mind any more than a simulation of the movement of molecules in a fluid should be expected to duplicate wetness, because the physical process underlying brains and computers are completely different.

I will argue that this idea is flawed.

Consciousness as an Emergent Property

Searle maintains that behaviour in itself is insufficient for consciousness. He says that consciousness is an emergent property of a system in which the right sorts of physical process are occurring. As an example of how this sort of reasoning works, imagine a large number of hydrogen and oxygen atoms, joined by covalent bonds to form water molecules, and a computer simulation of the interaction of a large number of hydrogen and oxygen atoms joined by covalent bonds. The simulation captures some of the *behaviour* of water, but the simulated water is not real water. *It is not really wet*. Searle would maintain that something like wetness is an emergent property of matter undergoing particular physical processes and that it is the same for consciousness: matter, undergoing the appropriate physical processes, has the emergent property of consciousness. Searle says that there is nothing mystical about this: it is just a fact about how physical systems behave. It is no more magical for consciousness to emerge from a system undergoing the relevant physical processes any more than it is for wetness to emerge from atoms undergoing the appropriate processes. Of course, you can get wetness in other ways, by using different atoms and making different molecules, but all of these ways will involve basically the same sort of physical processes. These physical processes are out of reach of the computer simulation and it will never cause real wetness. Likewise, a computer simulation of intelligent behaviour, in itself, is insufficient for consciousness. There may be ways of generating consciousness other than in human brains, using the same general kinds of physical processes. As we do not really know what these processes are yet, we cannot totally rule out a computer undergoing them and generating real consciousness, but Searle maintains that there is no reason to think consciousness is there, just because you have the right behaviour:

“...that is the old mistake enshrined in the Turing test. If this principle were correct, we would all have to conclude that radios are conscious because they exhibit intelligent verbal behaviour. But we do not draw any such conclusion.” [4]

To Searle, the underlying physics and processes are everything. You can think it probable that other people are conscious, not because they act as you do, but because you know that the physical workings of their brains are essentially the same: the same general kinds of physical effects occur in their heads as in yours, so the same emergent properties can be expected:

“Where knowledge of other minds is concerned, behaviour by itself is of no interest to us; it is rather the combination of behaviour with the knowledge of the causal underpinnings of that behaviour that form the basis of that knowledge.” [4]

The Problem

I disagree with Searle’s view that consciousness being an emergent property of physical systems makes it absurd to associate consciousness with the appropriate behaviour. My problem is that this idea seems to assume that there are two kinds of things we can observe in the world. One type of thing is a particular physical process, or the existence of particular physical objects; for example the interaction of water molecules and the existence of the real water molecules. I could produce two glasses of water and nobody would have problems with the idea that the same general physical processes are involved in each. The second type of thing is some kind of abstraction of a physical system in which the physical processes are not necessarily the same. For example, most people would admit that the same kind of *abstraction* could be made of a glass of water and a computer simulation of a glass of water, but would say that the underlying physical processes were different in each case. Underpinning the real wetness is the interaction of water molecules. Underpinning the simulated wetness is the interaction of electrons, which are not going to produce wetness and in any case are behaving in a way which appears, on casual inspection, totally different from the molecules in the glass of water. It is this kind of obvious analogy that makes Searle’s position seem so attractive: we know it is stupid to think that simulations of water molecules involve real wetness. My position is that this distinction between different kinds of properties is flawed. I am not saying that the simulated water is wet. What I am saying is that there is no profound difference between physical properties and processes and the sorts of abstractions of these things that are captured in computer models: everything is physical processes.

Suppose we want to know if some system is undergoing a particular kind of physical process. Whether I am right, or whether Searle is right, we should be able to formalize our approach, our rules that tell us when a particular physical process is happening. Suppose we do this by making a machine that I will call a “physical process detector”. The physical process detector works a bit like my “algorithm detector” in another article [10]. It has a set of probes with measurement tips that it can move around. A computer program controls the probes and the machine captures the measurements that it needs. The computer program analyzes the data and tells us whether or not the physical process(es) in which we are interested in is (or are) present. As emergent properties are caused by processes in a system then a physical process detector is also a physical property detector. We can use it to search for any emergent property that we wish, provided that we can define the processes associated with that property well enough to write the program. For example, we could program the machine to detect wetness and it would take readings to see if atoms were present and doing the sorts of things which cause wetness as an emergent property – using rules for “wetness detection” that we programmed into it. Maintaining that such a machine could not be programmed, in principle, to detect something like wetness would be maintaining that the concept of wetness could not be formalized, even in principle, and elevating it to a “supernatural”

status. I have argued in another article about the “supernatural” that this sort of idea would be incoherent [11].

If we properly programmed our machine to detect wetness and put it next to a glass of water it should detect wetness. If we put it next to a computer simulating the behaviour of water it should not detect wetness. (We will ignore any tiny amounts of water on the computer, for example from perspiration on the fingertips of people typing on it.): electrons are not wet and I will never persuade any of you that they are. However, suppose I defined a new type of property of a system and called it “b-wetness”. We could program the physical process detector to detect b-wetness when placed near a glass of water *and* when placed in front of a computer that is simulating the behaviour of water. It does this because the definition of b-wetness fits with processes in both the glass of water and the matter in the computer.

I am not trying to claim here that b-wetness is really wetness and that I can somehow use that to prove that wetness exists in the computer. It is not and it does not. B-wetness is a different property of a system and I only gave it a similar kind of name to make this easier to follow. Any system that has wetness as a property will also have b-wetness, but many systems will have b-wetness and not wetness. I do not think I need to define b-wetness fully, here, for people to see that this is possible.

If b-wetness is not the same as wetness then what use is this? The point is that b-wetness is an emergent property of a physical system, just as wetness is. There is a difference between the properties of wetness and b-wetness – different programs are needed to detect them – but there is no profound difference in the sort of property that is being detected: they are both physical properties of a system that can be found by a physical process detector. Someone might say, “Wrong! All cases of wetness involve the same general physical processes while b-wetness might involve many different types of physical processes.” I would say this is wrong. Wetness involves a certain general classification of physical process – that described by the wetness detection program. B-wetness involves a general classification of physical process – that described by the b-wetness detection program. We might see differences between systems in which b-wetness is detected – for example, the difference between a glass of water and the behaviour of electrons in a computer, but we can also see differences between systems in which wetness is detected – for example, the difference between the movement of water molecules and the movement of ethanol molecules. There is no respect in which one type of property is “physical” while the other is not. If you can say that wetness results when “basically the same sort of physical processes are occurring in a system” then I can say b-wetness *also* results when basically the same sorts of physical processes are occurring in a system. The processes might intuitively *appear* profoundly different but that is our bias: the processes producing b-wetness in a glass of water and producing b-wetness in a computer simulation of water have a lot in common, that commonality being described by the b-wetness detection program.

To this, someone could reply that “I don’t get it” and that the processes involved in wetness are really the same, while the processes involved in b-wetness are different, but

any such distinction is artificial. The wetness detection program would detect wetness in a large number of different systems, such as the same glass of water at different times or places, different glasses of water, glasses or bottles of different substances, fluids spilled on tables, etc. The behaviour of matter that produces wetness has a particular description which applies to a set of different systems. The behaviour of matter that produces b-wetness also has a particular description which applies to a set of different systems. The set of systems which have b-wetness will be much larger than the set of systems which have wetness, but that is just a matter of degree. The scope of b-wetness's definition is clearly much wider than that of wetness. We might argue that b-wetness is somehow more abstract than wetness, or more removed from the basic physical world than wetness, but this does not in itself mean that b-wetness is not an emergent property of a system. If being applicable to lots of systems means that something is not an emergent property then it could just as easily be argued that wetness is not an emergent property because it is claimed to be possessed by completely different physical systems. The formal description of any property will have a kind of "bandwidth" of systems to which it applies and it is just more obvious for b-wetness.

Someone might insist that I do not understand that wetness obviously involves the same physical processes while b-wetness is an abstraction. In reply I would suggest that we arrange all the possible detection programs in a row. On the left-hand side we have the programs that have "narrow bandwidth" – the ones that are very specific about what criteria have to be met for a property to be detected. The wetness detection program will not be on the extreme left because it has to have at least sufficient bandwidth to detect wetness in different glasses of water. The detection programs on the left-hand side will be much more specific than this – only detecting the property in a singular situation. As we go to the right the definitions of properties become more general, the corresponding detection programs detecting properties in larger sets of situations. As we keep going to the right, eventually we come to the detection program for the wetness detection property and if we keep going we encounter programs more general than the wetness detection program, eventually encountering the b-wetness detection program. As we go further we encounter still more general programs. The point is that *as we go from left to right there is no point at which things suddenly change and at which the detection programs stop detecting things that are not emergent properties of a physical system*. Wetness and b-wetness are just different emergent properties with different levels of generality. If you disagree with this my challenge to you is to tell me *exactly* at what point, as we go from left to right, the properties being detected suddenly become non-physical and "abstract", "artificial" or "made-up".

"You don't get it do you? Computers and brains do not involve the same processes."

Some readers may still think it obvious that a glass of water and a glass of ethanol involve the same general kind of process – movement of molecules – while a glass of real water and a computer simulation of water involve different processes – the movement of water molecules in a glass and the movement of electrons, in a different way, inside the computer. I am not doing something as naïve here as deluding myself by making some

crude mechanical analogy between the movement of electrons and the movement of water molecules. The only thing that justifies us in saying that a glass of water and a computer simulation of water involve the same physical processes is that the description of those processes could be expressed as a physical process detection program and used to detect b-wetness both in the glass of water and in the behaviour of matter in the computer. If this seems contrived it is for this reason:

There is a one-to-one correspondence between what is happening in a glass of water and a glass of ethanol. There is a simple one-to-one relationship between elements of the processes in the glass of water and the glass of ethanol: a single ethanol molecule is equivalent to a single water molecule, even if the behaviour is not exactly the same. The general description of wetness can be expressed as “X does various things” (let us assume those things are stated) and to get the description of what is going on in the glass of water we substitute “water molecules” for X and in the case of the glass of ethanol we substitute “ethanol molecules” for X. This one-to-one correspondence does not exist for b-wetness. Instead the general description of the b-wetness, expressed as a physical process detection program would have to work in a more complicated way to “find” b-wetness in both a glass of water and a computer simulating water. There is, however, nothing special about such one-to-one correspondence: it merely relates the description of the processes particularly directly to the system. It may seem “obvious” to us that physical processes defined so that there is such one-to-one correspondence to systems are “physical” because we are able to understand such processes easily, but to an entity much more intelligent than us much more abstract processes, in which the relationship between the description of the process and the elements of the system is very abstract, may appear trivially simple and “obvious”.

How this Relates to Searle’s Argument

Searle would maintain that there is no reason to think that a computer is conscious, even if it is behaving in the correct way, because the physical processes involved in human brains and computers are different. He also suggests that we can reasonably assume that other people’s brains have consciousness because the physical properties are essentially the same. The argument I have just made here should show that this distinction is artificial. Human brains may *seem* to have the same physical processes, and the same emergent properties, but to represent this with a process detection program in a physical process detector would still require the program to be general enough to detect consciousness in different human brains. We could define some emergent property in a general enough way that it is possessed both by AI systems and human brains and there would be no clearly profound reason why it was less valid. Just as Searle says that human brains involve the same physical processes, we could equally validly say that a computer running an AI program and a human brain involve the same physical processes if we define the physical processes in a general enough way – and you cannot validly argue that I am allowed to be general in my definitions of properties: you are doing that if you think other human brains are conscious by being similar, but not identical, to yours.

The implication of this is that if Searle can say that other human beings are probably conscious because their brains involve the same physical processes I could equally well say that I can expect certain computers to be conscious because, using more general descriptions of physical processes, they involve the same physical processes as my brain. Any profound nature of this argument is now lost. Searle may seem to be left with one advantage: the sorts of physical processes and properties needed to assume other brains to be conscious are less general than those needed to assume that brains and AI systems are conscious, but it is just a matter of degree. There is no sense in which this establishes that it is likely that other people's brains are conscious and unlikely that computers are not conscious. A third party could walk into such a debate and announce that only Searle's brain is conscious because it has the right physical processes and that neither my brain nor his (the third party's) qualify for consciousness because of the more general physical process detection program needed to include them.

Putting this back into the contexts of wetness and b-wetness, I could argue that consciousness is a very generally defined emergent property like b-wetness. Searle could argue that it is a more narrowly defined emergent property like wetness. A third party could say that we are both wrong and the processes needed for consciousness should be defined much more narrowly than either of us are doing. If Searle has any advantage by using more narrowly defined processes then someone who uses an even narrow definition would have an even greater advantage – leading to a solipsistic view of consciousness!

None of what I have said here implies that we need to presume that consciousness is not present when the relevant externally observable behaviour of a system is not observed. Searle correctly points out that there are medical conditions which can allow a human to remain conscious but not to show externally observable behaviour that we normally associate with consciousness. We could still presume consciousness to be associated with such systems from an inspection of their interior processes.

What about multiple realizability?

One of Searle's arguments involves what he calls "multiple realizability", and what I call "arbitrariness of interpretation" in some of my articles – the fact that any physical system can be said to be running any program by making the appropriate interpretation. That sort of issue can be raised for b-wetness. If we define a property so generally that a computer simulating water and a glass of water can have the same property then does this not mean that I can invent any physical property and say that it is shared by any set of systems, given some interpretation? For example, I could define some b-wetness property so that it is found in a glass of water, a block of concrete, a slice of dried out pizza and a cubic metre of vacuum. Does this not mean that trying to define properties like b-wetness is just making things up?

The problem with this is that *you can make the same objection against wetness*. Wetness may be defined in less general terms than b-wetness, but there is no obvious level of generality of the definition of a property at which properties abruptly become artificial. Multiple realizability is a problem that needs resolving. I suggest an approach in my

series of articles: *Minds, Substrate, Measure and Value* [7,8,9,10,11]. Considering this briefly, as it is not the main subject of this article, we might ask why, if presuming that an appropriately programmed computer is conscious because it works in the same general way as the human brain is valid, it is not similarly valid to say that a slice of pizza is conscious because it works in the same general way as the human brain: we could easily define a physical process detection program that finds the same property in a slice of pizza and a human brain, call it “consciousness” and demand human rights for pizza. I think this would be flawed though. I do not think it would make sense even to start considering an emergent property as a candidate for consciousness unless an abstract description of that property matched, in some way, our own mental experience. The emergent property would need to be a description, on some level, of what it seems like to be us. This would rule out lots of trivial emergent properties that we could easily find in brains and pizza. This does not solve everything, however: we could define an emergent property so that it is a candidate for consciousness and find it in both brains and pizza merely by basing it on a contrived physical process detection program. The important word here is “contrived”. The program needed to find this sort of property would be long and an argument could be made that this sort of program would have a “low measure” – a lower status, in some statistical sense – in the set of all physical process detection programs.

When we start to consider things in terms of measure, and how contrived our ways of finding properties need to be, this issue of multiple realizability starts to become tractable. It gives us an answer to Searle’s assertion that behaviour is of no interest at all [6,7,8,9,10]. Behaviour should be interesting to us because if two systems exhibit the same behaviour this suggests that it will be possible to write a relatively short and uncontrived physical process detection program that “finds” a suitable emergent property in each system. This does not mean that behaviour is everything: it may be possible to find a suitable emergent property, without a very contrived physical process detection program, in a system with no obvious, external evidence of intelligent behaviour. Behaviour, however, should be viewed as a strong indicator of consciousness because it suggests that extracting suitable properties will be easy.

It should be noted that this kind of justification for viewing computers as conscious is not the same as the standard strong AI justification and it is complicated by the issue of measure and it is possible to view this kind of situation with different levels of formality.

Minds and Reference Class

It may seem strange to talk about considering properties that could be candidates for our own mental states, but I would argue that that is what is done informally by science and humans in their everyday lives anyway. All that you can ever experience, by definition, is your own mental state. When you consider possibilities about how the world is it can only mean that, formally or informally, you are considering possible ways in which the emergent property of your mental state can have come about. This means that any scientific description of reality, as far as you are concerned, if properly expressed, should be a formal description of your abstract mental state and how it relates to the rest of

reality. As we do not know everything about our own mental states or reality we have some uncertainty about our mental situations. We have a reference class of possible mental states we could be in and for each of these a reference class of possible descriptions of how those mental states relate to reality. I would argue that the best way of approaching consciousness is to look at it in this way and ask what kinds of situations we should be prepared to admit *in principle* into our reference classes of possible situations in which our mental states could be. I explore this idea in another article [10]. When considered like this, questions such as “Can a computer be conscious?” become “Is it conceivable that a description of my mental state which involves it being an emergent property of the processes in a computer could ever in principle be part of my reference class of possible situations?” I suggest that such a reference class should include every possibility which involves some formally describable way of relating a valid description of a mental state to physical reality and that the possibility, in principle, of including computers in such a reference class suggests admitting them as reasonable candidates for consciousness, though possibly with some consideration of measure.

We can get an idea of what this reference class view would mean by considering various thought experiments. For example, suppose you wake one morning remembering that a mad scientist you met at a party had abducted you and was going to make you unconscious and scan your brain during the night, making a computer simulated copy which was to be woken the next morning in a virtual reality. Assume that you live in a world where the technology to do this is routinely available – and nothing that Searle says opposes such an idea; in fact it could be argued that his materialistic view of consciousness pretty much mandates it being possible. Situations similar (in some ways) to this are described in science fiction novels [12,13]. Are you the original person or are you the copy? If you approached this by having an automatic bias against being the copy on account of consciousness being unlikely in computers I would say you are taking a flawed position. You cannot even use an argument such as “I know human brains tend to be conscious because they have the sorts of processes causing my consciousness”. In fact, if Searle’s methodology is valid you cannot even be sure that human brains *are* conscious: you could be in the computer experiencing consciousness with memories imported from a brain that had never experienced consciousness: you might only be experiencing consciousness at all as a result of copying. You would need to assume you are not in a computer program to get Searle’s reasoning off the ground in the first place. The only way to approach this, without making artificial special cases and assumptions, is to consider the reference class of possible abstract descriptions of your mental states and, for each, the reference class of possible ways in which the relationship between the mental state and the rest of the world could be described – how the mental state could follow-on from everything else as an emergent property. In practice, the reference class of possible ways of relating a mental state to the rest of reality is probably not going to depend too much on the subtleties of the mental state and you can probably do almost as well by selecting one reasonably likely mental state and considering the reference class of possible, formally describable situations in which it could be. Considerations of measure should play a part in this kind of consideration, as discussed in my other series of articles [6,7,8,9,10]. The point is that such a reference class approach seems to be almost mandatory to make sense of this kind of situation, particularly if we start to consider

other issues that are discussed in these other articles, and admitting its use is very suggestive that a very generalized approach to consciousness, with consciousness described as an emergent property in the most general terms possible, is warranted.

Are the physical processes even *remotely* similar?

Some people may reject my assertion that physical, emergent properties can be considered at such a general level, or they may accept it, but say that even the matter of degree involved is hugely important because the physical processes in human brains are *almost exactly* the same, allowing a very narrow classification of physical process to be considered, while any physical processes that appropriately programmed computers and brains have in common, even if these are admitted, are much more different from each other, requiring a very wide classification of physical process to be considered.

Against this, I would first say that it is basically an appeal to incredulity. Another point I would make is that we cannot even be sure that the physical processes in different people's brains are even *remotely* similar. In fact, we cannot even be sure that the physical processes in two brain cells or even two carbon atoms are remotely similar.

How can this be the case? The processes may seem very similar from our human point of view, considering what is known about physics, but we do not know that that physics is fundamental. It could be that underpinning a "fundamental" particle there is a lot of physics, and a lot of processes, that we do not know about. If this happened to be the case we could not be sure that two "fundamental" particles (though of course they would only seem "fundamental" to us if this were the case) were there because of remotely similar processes. Suppose we choose some apparently fundamental particle of type "X". It may be that any particle of type X just happens to be a particularly stable and common type of emergent property that arises in many different ways, with many different types of processes underneath it and it may be that if we had to consider the (currently unknown) lower level physics as well, then the description of the general type of processes underpinning X particles needed to find them, using the physical process detector discussed previously, lacks any simple one-to-one relationships and is vastly more complex, more abstract and, to us, apparently more contrived than the sorts of programs needed to "find" similar physical properties in brains and appropriately programmed computers.

If this were the case it would cause problems for Searle's position. How could we say that two brains use the same general kind of physical processes when it was not even possible to say that two "fundamental" particles in the two brains, or even in the same brain, existed according to the same physical processes? We could say that the processes underpinning the particles do not matter and that the appearance of "fundamental" particles as being the same means that we can ignore what is underneath, but that seems to be specifically what Searle wants to prohibit us from doing on a larger scale with computers and brains. If process is everything, and behaviour is nothing, then how could we say that one particle of type X was the same sort of emergent property as another particle of X if only the behaviour were similar? Another way out of this problem may

seem to be by saying that whatever causes two particles of type X must be the same sort of phenomena, even if how it occurs is different. For example, suppose the physics underpinning particles of type X is the “zod soup”. (I am just inventing some nonsense physics terms here to avoid having to attack or defend any specific views in physics.) It might seem that we can declare all X particles to rely on the same processes by virtue of all being based on “zod soup” physics, but just being based on the same physical laws would hardly justify this. It may be that the description of what is going on in the “zod soup” to get one X particle is vastly different from what is going on in the “zod soup” to get another X particle and even if the two kinds of processes can be made to sound the same with generalizations, we could do the same with computers and brains: I doubt, for example, if Searle would be impressed if I said that computers and brains have the same causation for their behaviour because they both use condensed matter. It is hard to escape this: if physics at a low level happened to mean that two X particles, if we applied Searle’s standard, did not share the same causation then it would be invalid to ignore this and declare the same causation to exist in things made of X particles.

I am not saying that physics is like this, but I am saying that Searle’s argument rests on the assumption that it is not. It does not end with particle physics. Even if the underpinnings of particle physics “cooperate” with Searle, we could ask the same questions about what, if anything underpins the physics underpinning particle physics. Searle’s argument depends on physics not turning out like this *all the way down*.

I would be interested in knowing if Searle accepts the possibility, even in principle, that physics could turn out like this. If he does, I would also be interested in knowing what he thinks the implications would be for his argument if physics did turn out to be like this.

Are you really saying that there is no such thing as abstraction?

You might consider what I have said here in a more general way than just with regard to Searle’s argument. It might seem that I am arguing that everything is physical and that abstraction does not exist. In a way, I am. How we describe this, however, depends on semantics. Stated one way, I could be viewed as saying that there is no such thing as “abstraction” and that everything is physical. Stated another way I could be viewed as saying that everything is abstraction and that the “physical” is just a special case of abstraction, special only in that it appears at a particularly low level of nature.

This may be hard to accept. It may be “obvious”, for example, that an abstract concept such as “money” is somehow *qualitatively* different to heat. Heat seems to relate only to the physics world whereas “money” seems to depend on the subjectivity of human minds. Heat, however, only exists because of the particular way in which matter behaves. Human minds are associated with brains, which are also made of matter. While the concept of “money” may be viewed as “subjective” to us, to some entity outside of humanity, and outside our ideas of economics it would just be viewed as an emergent property of societies, which are just an emergent property of groups of minds, which are just an emergent property of single minds, which are just an emergent property of brains,

which are just an emergent property of molecules and so on. Considering “money” as “subjective” merely means that when brains change, the *physical*, economic properties change. The appearance of such “subjectivity”, if we regard it as a profound idea, only comes about because we award ourselves a special place in the taxonomy of things. This even applies for concepts such as “beautiful”. It may seem that this is somehow a “non-physical” property, but it is an emergent property of matter, just as your mind or your brain are. The fact that you can have one idea of “beautiful” and someone else can have another merely means that two different properties are involved which should ideally be given different names. Human language is not constructed for philosophical utility and we do not give separate names to our different ideas of “beautiful”. This may lead us to think that the “subjectivity” of the concept somehow makes it “non-physical”. We could even understand “beautiful” as a property without having to assume a separate concept of “beautiful” for everyone. If we had the time we should be able to construct some general, over-arching description of what is going on when something is being considered beautiful and which applies to all brains: your particular concept of “beautiful” would be a special case of this.

Conclusion

I have argued against Searle’s view that it makes no sense to say that we can presume that other humans are conscious because of similar physical processes in their brains, while dismissing a suitably programmed computer as a likely candidate for consciousness. My argument is based on the idea that saying that two systems have the same physical underpinnings is subjective and that this becomes obvious when we try to formalize our idea of what it means to say that two systems are exhibiting the same properties. Searle regards behaviour in itself as irrelevant for deciding if a system is conscious, but we should view behaviour as important because it is a strong indicator of how contrived a description of some emergent property would need to be so that that property is a candidate for consciousness and the system shares it with human brains.

Even if I am right, there is still a debate to be had here. The issue of measure complicates matters, though I discuss this in another series of articles [6,7,8,9,10]. None of this proves that computers are conscious, nor does it refute Searle’s Chinese room argument [5], or prove that strong AI [1] is a coherent, well-defined or valid position. What I have hopefully done is give some clarification to the particular part of this debate that is about the “difference” between simulated and real processes.

References

- [1] Searle, J. R. (1980). Minds, brains and computers. *The Behavioral and Brain Sciences* 3:417-457.
- [2] Searle, J. R. (1997). *The Mystery of Consciousness*. New York: The New York Review of Books.

[3] Searle, J. R. (2002). *The Rediscovery of the Mind*. Cambridge, Massachusetts: MIT Press. (9th Printing, Originally Published: 1994. Cambridge, Massachusetts: MIT Press).

[4] Ibid. Chapter 1, pp22.

[5] Ibid. Chapter 2, pp45.

[6] Web Reference: Almond, P. (2007). *Minds, Substrate, Measure and Value, Part 1: Substrate Dependence*. Retrieved 12 September 2007 from <http://www.paul-almond.com/Substrate1.pdf>. (Also at <http://www.paul-almond.com/Substrate1.htm>).

[7] Web Reference: Almond, P. (2007). *Minds, Substrate, Measure and Value, Part 1: Substrate Dependence*. Retrieved 13 September 2007 from <http://www.machineslikeus.com/cms/minds-substrate-measure-and-value-part-1-substrate-dependence.html>. (A copy of the article in Reference [7]. Includes reader criticism of the article).

[8] Web Reference: Almond, P. (2007). *Minds, Substrate, Measure and Value, Part 2: Extra Information About Substrate Dependence*. Retrieved 3 November 2007 from <http://www.paul-almond.com/Substrate2.pdf>. (Also at <http://www.paul-almond.com/Substrate2.htm>).

[9] Web Reference: Almond, P. (2007). *Minds, Substrate, Measure and Value, Part 2: Extra Information About Substrate Dependence*. Retrieved 10 November 2007 from <http://www.machineslikeus.com/cms/extra-information-about-substrate-dependence.html>. (A copy of the article in Reference [9]. Includes reader criticism of the article).

[10] Web Reference: Almond, P. (2008). *Minds, Substrate, Measure and Value, Part 3: The Problem of Arbitrariness of Interpretation*. Retrieved 11 May 2008 from <http://www.paul-almond.com/Substrate3.pdf>. (Also at <http://www.paul-almond.com/Substrate3.htm>).

[11] Web Reference: Almond, P. (2008). *Against the Supernatural as a Profound Idea*. Retrieved 1 November 2008 from <http://www.paul-almond.com/Supernatural.pdf>. (Also at <http://www.paul-almond.com/Supernatural.htm>).

[12] Egan, G. (1994). *Permutation City*. London: Millennium. (Fiction).

[13] Ballantyne, T. (2004). *Recursion*. London: Tor UK. (Fiction).