Why physicalism seems to be (and is) incompatible with intentionality

Richard Johns
Dept. of Philosophy
Langara College
rjohns@langara.bc.ca
May 2013

Physicalism, and the mechanical philosophy that preceded it, face a common criticism: we have a strong intuition that nothing fully physical or mechanical could possibly be a mind. In this paper I will investigate the source of this intuition, and use a new argument to show that it is in fact correct.

The feature of physical and mechanical systems that drives this intuition, I will argue, is that such systems have states that are entirely clear and intelligible, having precise mathematical characterisations. We shall therefore begin by recounting the historical origins of this idea, and its role in arguments against physicalism.

1. Mechanism, Physicalism, and the Mind

The pioneers of the mechanical philosophy, philosophers such as Boyle, Descartes and Gassendi, saw no prospect of understanding everything in mechanical terms. Some tricky properties, such as colours, were declared to exist only in the mind of the observer. Of course that merely moved the problem rather than solving it, as one must still give an account of the mind. (But many followed Descartes in seeing the mind as essentially non-mechanical.)

Nowadays the strict mechanical philosophy, where the world consists of particles with only geometrical qualities, is not tenable. Even in Newton's day the gravitational force was problematic, as it appeared to act magically at a distance without any contact between the particles affected. More recently physicists have discovered that the physical world is more

subtle than Descartes could have imagined. Fields are needed, as well as particles, and the mathematics needed to describe such fields is highly abstract, and very remote from Euclidean geometry. Thus today there are no mechanists, of the 17th century sort. Nevertheless the spirit of mechanism (so to speak) is still alive and well in the 21st century, and goes by the name 'physicalism'. Quantised fields may not be mechanical, but they are still fully physical.

Physicalism is not nearly so sharply defined a doctrine as mechanism was, however. If the world isn't a collection of particles, then what is it, exactly? There is unfortunately no generally-accepted definition of physicalism, but in this paper I shall assume that physical systems are all *transparent*, which here roughly means that they can be perfectly comprehended by minds of sufficient ability (especially mathematical ability). The notion of transparency is discussed in more detail in the next section.

It is remarkable that physicalism, despite its great differences from the mechanical philosophy, is also attacked for an alleged inability to account for the mind. In my view this important similarity between mechanism and physicalism results from the fact that both assert the complete transparency of the world. I believe that it is the transparency of physical states that grounds the intuition (whether correct or not) that intentional states and phenomenal experiences are more than "merely" physical.

2. Physicalism and Transparent Qualities

The pioneers of the mechanical philosophy had no grounds to think that mechanical qualities would be adequate to explain all phenomena. They nevertheless urged that mechanical explanations be sought, where possible, on account of their clarity. Boyle (1674), for example, begins his defense of mechanism by saying, "And the first thing that recommends it is the clearness and intelligibleness of its principles and explanations". He contrasts corpuscular properties with those used by the Aristotelians and chymists, since the latter are 'dark', 'obscure', 'occult', etc.

Transparency, in the sense used here, is the same as Boyle's "clearness and intelligibleness". But what does it really amount to? Let us begin by asking why mechanical explanations were so clear and intelligible. Such explanations viewed a system as consisting of particles whose properties were just shape, size and motion – geometrical (and kinematic) properties, in other words. Now geometrical qualities – square, triangle, plane, sphere, etc. – are intelligible in the sense that that are, or exactly correspond to, mental concepts. For Plato these geometrical Forms were accessible only through the mind, not the senses. Being accessible to the mind entails that geometrical qualities like *Square* have no occult qualities, ones that are hidden or beyond our ken. When we define a square (in the usual way) we specify the concept entirely, leaving nothing vague or indeterminate.

Whether or not Plato was correct in his realism concerning geometrical and mathematical entities, we can agree with him that these entities are fully intelligible. (If such entities are mere creatures of the mind, for example, then this must also be true.) So we see that the transparency of mechanical systems derives from the intelligibility of mathematical objects.

In that case, however, physical systems are also transparent in exactly the same way, and to the same degree. For, as noted above, when physics gave up on a mechanical understanding of the world, it retreated to other forms of mathematical description. The mechanical models of electromagnetic fields that Maxwell devised were later considered superfluous to a physical understanding—as Hertz famously put it, "Maxwell's theory is Maxwell's equations". Furthermore, no present physical theory (that I am aware of) claims the existence of any property that cannot be mathematically characterized in some way. How would present physicists respond to claims that a certain property of the electron is 'occult', or not amenable to mathematical treatment? They would dismiss it as not part of physics, at least, on the grounds that one could not calculate any observational consequences from the theory.

It is true that today 'physicalism' today has a variety of meanings, and not all are committed in this way to the full transparency of physical states. According to Stoljar (2001: 257) for example, a rock is going to count as a physical object (an 'o-physical' object) no matter

what its ultimate nature turns out to be. Physicalism, as defined in such a broad fashion, clearly is not refuted by anything in this paper. One may question, however, whether such views deserve to be called 'physicalism', for they seem to have no connection to physics, or to the mechanical philosophy that physicalism succeeded.

Finally, it must be stressed that while the simple transparent properties discussed so far are all ones that humans can grasp clearly, this need not be true of transparent qualities in general. Since human minds are finite and limited, there may be transparent properties that are simply too complex for us to grasp. Such properties are very different from occult qualities, however, since they are fully intelligible in principle, by a being such as Laplace's demon, just not by us.

3. Transparency and the Mind

As stated above, I claim that the apparent inability of mechanical (and physical) explanations to account for the mind results from the *transparency* of mechanical (and physical) processes. The role of transparency in this intuition seems to be crucial to much, and perhaps all, opposition to mechanism and physicalism.

Blaise Pascal (1669) states, for example, "... if we were simply material, that would prevent us from knowing anything at all, there being nothing so inconceivable as the idea that matter knows itself. We cannot possibly know how it would know itself."[1] The argument here, as brief as it is, involves *our* inability to know how matter could have knowledge. Now why should our inability to know some proposition *p* entail that *p* is false? Pascal is apparently assuming that 'matter' is transparent, or intelligible, so that its properties are readily apparent to

[1]From Pensée no. 72. Original text: "... quand on prétendrait que nous serions simplement corporels, cela nous exclurait bien davantage de la connaissance des choses, n'y ayant rien de plus inconcevable que de dire que la matière se connaît soi-même; il ne nous est pas possible de connaître comment elle se connaîtrait."

us. If a transparent substance were to "know itself", then we humans should be able to see how it knows itself. (Also, Pascal seems to take it for granted that no such knowledge will be forthcoming.)

Later, in the 18th century, Leibniz (1714, Sec. 17) argued in a similar way that "Perception, and that which depends upon it, are inexplicable by mechanical causes, that is to say, by figures and motions." To support this claim Leibniz imagines enlarging an alleged thinking machine to the point that one could walk around inside it, as in a mill. In that case the workings of the machine would be fully transparent and open to inspection, yet such an understanding of the machine would not reveal any thought at all, but only "pieces working upon one another".

The same assumption of the transparency of physical states is also at work in 20th century arguments against physicalism. In Frank Jackson's knowledge argument, for example, we are invited to consider the case of Mary, a neuroscientist who has never had a colour experience, yet who has complete physical information about the physiology of human colour vision. Jackson clearly supposes that physical qualities are fully intelligible, as he states that, according to physicalism, Mary ought to know *what it is like* to have (say) a red colour experience. If the physiology of vision involved occult qualities of some kind, opaque to the intellect, then how could Mary make an inference of this sort?

Jackson's thought experiment is intended to pump our intuition that the known properties of matter are logically independent of properties of conscious experience. (David Chalmers (1996) describes this as an 'epistemic gap'.) One cannot, for example, define the colour red in terms of geometrical properties, or any other mathematical properties. Thus the overall scheme of arguments such as Pascal's, Leibniz's and Jackson's may be summarized as:

1. The intelligible properties of physical systems are logically independent of mental properties like colour experiences.
2. Physical systems are transparent, so that all their properties are intelligible.

∴ Physical systems do not have mental properties

My argument below is rather different to this pattern, in that I do not use premise 1. Also, my argument is focused on intentionality rather than phenomenal experiences.

4. A Useful Analogue

Before giving my reductio of physicalism, I will illustrate the general kind of argument I'm using by proving the rather trivial result that, for any formal language of arithmetic, some properties of natural numbers are not expressed by any well-formed formula, or 'wff' (with one free variable) of that language.

Take a formal language of arithmetic (call it *L*). Gödel showed that the wffs of a formal language can be assigned unique code numbers, and so suppose that such an encoding has been chosen for *L*. Then, for any given wff of *L* (with one free variable), it's always meaningful to ask whether it is satisfied by its own Gödel number. For example, if all the code numbers of wffs are even numbers, then the wff that expresses '*x* is an even number' will be satisfied by its own code number. We can then define a natural number *p* to be *inclusive* (w.r.t. *L*) just in case: (i) *p* is the code number of a wff Φ with one free variable, and (ii) *p* satisfies Φ. It is readily apparent that, while *inclusive* is a well-defined property of natural numbers, it is not expressed by any wff of *L*. For suppose there is such a wff, say Ψ. Then, since any wff can be negated, there is also a wff ¬Ψ, expressing the property of being non-inclusive. The wff ¬Ψ has a code number, say *g*. We then see, of course, that *g* satisfies ¬Ψ if and only if it does not, which is a contradiction.

While the argument is simple, it is strongly analogous to the main argument below and helps to clarify some important points. The first is that the structure of the argument is exactly like the Grelling-Nelson 'heterological' paradox, where a word is 'heterological' just in case it doesn't apply to itself. (E.g. 'long' is a heterological word, since it is a short word, not a long one.) When we ask whether 'heterological' is itself heterological, we obtain a contradiction, since it is heterological if and only if it is not heterological.

The usual response to this problem, which I accept, is to say that the scope of the term 'heterological', as provided by its definition, does not legitimately extend to the word 'heterological' itself. I am not sure exactly how far the definition extends, but it can safely be applied to adjectives that express 'straightforward syntactical properties' of words, such as long, short, four-lettered, pentasyllabic, and so on. On the other hand, if the adjective expresses some property with a *semantic* element, as 'heterological' does, then we might find that it falls outside the scope of the definition.

The above proof about formal languages is not paradoxical. It skirts the Grelling paradox, just as Gödel's proof of the first incompleteness theorem skirts the liar paradox. How does it avoid paradox, given the close formal similarity?

The big difference is that the arithmetical case starts with the wffs of a formal language of arithmetic. We know that all of these wffs of *L* have clear and precise meanings, as they are just talking about numbers, in a straightforward way, using + and × as well as the logical operators. Each one is satisfied by a definite set of natural numbers. Not one of these wffs is vague, ambiguous or paradoxical. Thus, *even though 'inclusive' is defined semantically*, using the notion of satisfaction, we know that it is well defined over the entire set of natural numbers. The second difference is that the contradiction in the formal case arises only when we assume that 'inclusive' is expressed by some wff of *L*. This assumption, which has no counterpart in the Grelling-Nelson paradox, is therefore refuted when the contradiction is derived.

After presenting my simple refutation of physicalism I will show that, like the arithmetical argument above, it avoids paradox.

5. The Main Argument

The notion of a physical system with intentional states (such as beliefs) is implicitly contradictory, I will now argue. The argument depends on the premise that physical systems are entirely transparent to the mind, being fully characterized by mathematical structures, or other devices of similar clarity.

The construction below involves, as one might expect, some self-reference. If a physical system (I'll often say 'machine' for brevity) can think, and its own states are transparent, then it can have thoughts about its own states. This allows the construction of a diagonal property, one that the machine cannot use to form thoughts, on pain of contradiction. The diagonal property in question is analogous to the term 'inclusive' used in the argument above about formal arithmetic, and is easy for a human logician to understand.

This main argument is of the form *reductio ad absurdum*, and the reductio assumption is that there exists a physical system ('machine' for short) that literally has beliefs. Like many science fiction characters, this machine is assumed to have an intellect roughly equal to that of an intelligent human. The argument also assumes that these beliefs logically supervene on the complete physical state of the machine. Thus Laplace's demon (a theoretical being of unlimited cognitive ability) could rationally infer the content of the machine's beliefs, at a given time, from a sufficient physical description of it at that time. We do not assume, however, that the machine itself can completely understand its own states. A partial understanding, of some aspects of its own workings, will suffice.

Given these assumptions, there will be some class C of possible physical states of the machine. Also, for each proposition *P* that the machine is capable of conceiving, there will be some subset of states in C in which the machine is thinking about (i.e. at least reflecting on or considering) *P*. For convenience, I will assume that the machine is capable of consciously considering only one proposition at a time, so that each physical state is associated with one proposition at most. In that case, a state in which a proposition *P* is being considered is

effectively a sentence that expresses *P*, in the sense that it is a physical object that precisely specifies the proposition *P*. Moreover, the subset of states in C that express propositions in this way is effectively a language, so we will call this subset *L*. Note that the 'language' *L* is rather unusual in that it needs no separate semantics. The semantic content of each 'sentence' of *L* is a logical consequence of its physical properties.

Some care is needed when using the term 'physical state'. I assume here that for each possible physical state there is a corresponding physical description that uses the standard terms of physics and mathematics. For example, the state of a point particle in classical physics corresponds to a description of its position and momentum, as mathematical vectors. Thus, even for the states that are also 'sentences' of *L*, and hence semantic entities in some sense, there is a complete description in the ordinary language used by physicists.

Moreover, following standard terminology, I assume that every state that logically supervenes on purely physical properties is itself purely physical. In this way, chemical and biological properties will also count as physical properties, just so long as Laplace's demon could infer these properties for an object, given a complete physical description at the fundamental level.

Among all the sentences of *L*, we will focus on just those that express existential propositions, i.e. ones of the form "at least one thing has the property *P*", where *P* is any physical property, i.e. any property that can be expressed in ordinary physical terms (or logically supervenes on such properties). It is then possible that there are physical states S in *L* such that, when the machine is in state S it is considering the proposition "something has the property *P*", and the state S itself has the property *P*. For example, if the machine is a digital computer made of logic gates, the machine might be in a state S where it is considering "Logic gate #76352 is off in at least one state", and logic gate #76352 is off in that very state S itself.

A state of this sort is analogous to an 'inclusive' natural number as previously defined, so we shall call it an *inclusive state*. Note that, to count as an inclusive state, the property *P* involved in the existential thought must be a *physical* property, i.e. one that can be expressed in

physical terms, although the same property might be definable in other ways, perhaps indirectly or relationally.

With 'inclusive' thus defined, we need to ask whether 'inclusive' itself is a physical property. It may appear that it is not, since it is defined using the semantic notion of an object having a property, but we should remember that the same property can be defined in different ways. In fact it is easily shown that 'inclusive' is a physical property, as follows. Suppose that Laplace's demon has complete physical information concerning a state S in *L*. In that case, the demon knows what the machine is thinking about in state S, i.e. it knows what property *P* appears in that existential belief. Since *P* is itself a physical property (by definition), the demon knows whether or not S has the property *P*. In other words, based on the purely physical description of the state S, the demon can infer whether or not S is inclusive. 'Inclusive' is a physical property by virtue of logically supervening on physical properties.

On the other hand, it is also clear that 'inclusive' cannot be a physical property. For, since we have been able to define it here, in a rigorous way, the thinking machine would be able to understand the concept of an inclusive state and form thoughts about it. In particular, the machine might then consider the proposition "Some states are non-inclusive". If the machine is able to think this thought, however, then there is some state S' in which it does think it. The existence of such a state S' is impossible, however, since S' would be inclusive just in case it is non-inclusive.

Our reductio assumption of a thinking machine entails a contradiction, and so must be rejected.

6. Comparison with Gödel, Lucas and Penrose.

My argument is similar in some respects to one based on Gödel's first incompleteness theorem, which has been made by J. R. Lucas (1961), Roger Penrose (1989, 1994), and Gödel himself. In particular, both involve convoluted diagonal reasoning. (I will refer to this as the Gödel

argument, since he was apparently the first to make something like it in his 1951 Gibbs lecture.) I will briefly summarise this type of argument, in order to clarify some differences between it and my argument of Section 5.

The Gödel argument exists in various versions, but the general idea involves a comparison between the mathematical sentences provable in a formal system and the mathematical theorems that can (in principle) be proved by human mathematicians. Human mathematicians seem to be equivalent to one another, in the sense that the theorems proved by one mathematician are generally accepted by others as valid. Where differences of opinion arise, they are usually resolved by friendly discussion, to everyone's satisfaction. (The mathematicians who change their minds in these discussions do so because they 'see for themselves' that the others are right.) One might then assume that there is some formal system F whose axioms and rules of inference prove the same set of theorems that human mathematicians can (in principle), but the Gödel argument concludes otherwise. For Gödel showed that any such system F will have a "Gödel sentence", which effectively asserts its own unprovability in F. By understanding the meaning of the sentence in those terms, a human can infer the truth of the Gödel sentence from the soundness of F, since the falsehood of the sentence would entail its provability in F, which is a contradiction. Since the sentence is true it must therefore be unprovable in F, since that is just what it claims. Now suppose that humans are legitimately convinced that human mathematical reasoning is sound. In that case, a human who constructs the Gödel sentence for F can also prove that it is true, which the system F itself cannot do. It then follows that F doesn't capture human mathematical knowledge at all. But in that case, since this argument may be used for any system F whatsoever, it follows that human thought isn't formalisable at all.

My argument differs from the Gödel argument in a number of important respects. For example, my argument has nothing to do with formal systems, and does not assume for example that human mathematical reasoning is sound. These differences are rather obvious, and entail that few if any of the usual objections to the Gödel argument can be applied to mine.

My argument differs from the Gödel argument in another respect that is more subtle, however. The Gödel argument depends on the claim that, for any given system F, a human can construct and prove the particular Gödel sentence for F. (Note that, while there is a Gödel sentence for each system, there is no general one that applies to all systems.) Thus the argument assumes that humans can fully comprehend formal systems of arbitrary complexity. This opens the argument to the objection that there is a hierarchy of formal systems, each able to prove the Gödel sentences of 'lower' (simpler) systems, but not its own Gödel sentence, or those of higher systems. Human mathematics is equivalent to a formal system somewhere in this hierarchy, so that if some superior being did present us with our own formal system, we would find ourselves unable to understand it well enough to construct our own Gödel sentence.

My argument, in contrast, does define a single property, 'inclusive', that depends only on the general idea of physicalism, and not on any details of any specific physical system. In assuming that an intentional physical system would understand this term we do *not* assume it can fully grasp its own physical states. It is enough that the machine have some concepts that some of its physical states fall under. As Pascal (nearly) said, it is impossible for us to think that a physical system can be aware that it is a physical system.

7. First Objection: Just another semantic paradox

The main objection to the argument of the previous section is likely to be one that I have already mentioned, that the contradiction obtained is just another semantic paradox. In that case, the contradiction probably has nothing at all to do with physicalism, but is due merely to whatever ordinarily creates a semantic paradox (insert your favourite theory here).

To evaluate this worry I invite the reader to compare my argument to the Grelling-Nelson paradox, and to the proof about arithmetical wffs in Section 4. The definition of 'heterological' is surely sound as long as its scope is restricted to purely 'physical', or syntactic, features of words. It is only when it is applied to a semantic term like 'heterological' (or its contrary,

'autological') that one has the nasty feeling of a closed loop. It is for this reason that, when defining 'inclusive', the scope of the term is restricted to purely physical properties. The definition of 'inclusive' therefore, is not problematic in the way that the definition of 'heterological' is.

The heart of my reasoning is the argument that 'inclusive' is a physical property. That argument has no counterpart in the derivation of the Grelling-Nelson paradox. It does however correspond to the proof about arithmetical wffs, in that 'inclusive' as defined there is a well-defined property of natural numbers, one that could presumably be defined in some more straightforward way, though not in the language *L*. The reason why 'inclusive' is provably a physical property is that the physical states in *L* logically entail their own meanings. This premise, which is a consequence of physicalism, is crucial to the argument.

8. Second Objection: The argument proves too much

The contradiction derived in Section 5 assumes nothing about physical states except their being transparent to thought. It may then be objected that the argument, if correct, would refute not only physicalism but almost every conceivable theory of the mind. In particular, it would even refute theism, since surely the human mind is transparent to God at least. After all, whatever the ultimate properties that are needed for a being to have consciousness, intentionality, etc., surely the world's creator can understand them?

A quick response to this is that, even if the argument did apparently prove too much, critics should not consider the argument refuted on this basis alone, without finding a flaw in the argument itself. For example, if someone showed me an apparently cogent argument that 2+2=5, I should be quite sure that there was a mistake somewhere, yet I would not rest easy until I found the mistake.

A proper response to this objection, which is sketched below, should show that the transparency assumption is denied by some theories of the mind, and that these theories are none

the worse for it. Indeed, I shall argue that such theories have certain advantages in addition to avoiding the contradiction of Section 5. Nevertheless, I am not arguing for these accounts here, which would require a paper of its own.

A good first step to see how the transparency assumption can be rejected is to consider some comments of Bertrand Russell (1927) about physics. Russell's view is that, while physics presents the natural world as being purely mathematical, this is not a complete picture of reality. Rather, this is just the only aspect of reality that we can discover using the tools of physics. The natural world has other aspects, such as phenomenal experiences, which cannot be described mathematically. Other philosophers[2] have since argued similarly that physics describes only the relational properties of matter (omitting the intrinsic properties) and only the dispositional properties (omitting the categorical properties).

This view has a rather Kantian sound to it, in claiming the existence of a 'noumenal' world that is beyond our ken, but it need not be a version of idealism. The mathematical qualities that we *can* understand should, I believe, be considered to exist in the world itself, independently of us. The claim that our understanding of reality is incomplete – even permanently so – does not, by itself, lead to any form of anti-realism.

In the remainder of this section I will sketch a type of theory of the mind, inspired by Russell, that might be called 'concretist', for lack of an existing term. It will be clear that concretist theories are incompatible with the view that mental states are transparent, as physicalism claims. I hope that concretism will also seem attractive on its own terms.

A concretist view about the mind says that the mind is essentially a concrete object, and cannot be fully understand in abstract terms, such as geometrical or mathematical terms. A concretist will say, therefore, that the mind cannot be fully grasped by *any* mind at all. It is not fully transparent, in the way that a mechanical system is. One can also be a concretist about other topics, such as causation and the flow of time.

---

[2] See for example Lockwood (1991), Stoljar (2001), and Strawson (2008).

Concretism depends on there being a significant difference between abstract and concrete objects, where abstract objects are those that are 'objects of thought', being perhaps products of rational thought, or at least objects that are fully (and only) accessible to thought. Some common examples of abstract objects are numbers, sets, propositions, properties, and possible states of affairs. Abstract objects exactly correspond to ideas, and so speaking very loosely we may say that abstract objects are ideas. Concrete objects, by contrast, are not ideas. They are accessed through the senses, via their causal interaction with us, and are not fully intelligible on this view.

The difference between abstract and concrete objects on this view is a matter of what is often called 'real existence'. Consider for example a physical system whose behaviour must satisfy some equation of motion. In that case, each solution to the equation represents a possible history of the system. One of these solutions actually takes place, in the sense that it represents the actual motion of the system – it 'really happens'. Now, what quality of this actual history distinguishes it from the myriad of possible histories? What extra ingredient does it have? Two things are obvious here. First, that this quality of concreteness or 'real existence' is not something that can be expressed mathematically. After all, that very solution might not have been the real one, and in that case its mathematical properties would have been exactly as they are. Second, physics as a subject has nothing to say about real existence, in the sense that physicists don't write papers about it, or construct theories of it. (Though physicists sometimes muse about the puzzle of real existence in an occasional philosophical moment.[3])

Since the mathematical solution to the equation of motion would have existed abstractly, even if the real motion were otherwise, I think it is helpful to distinguish that solution from the real (concrete) history. The actual solution is somehow 'part of' the real history, since the real history has the properties that the actual solution describes, but there is some extra element in the real history, whose old-fashioned name is *substance* or *substratum*, meaning 'standing under' or

---

[3] For example, Stephen Hawking (1988, p. 192) wonders: "What is it that breathes fire into the equations and makes a universe for them to describe? ... Why does the universe go to all the bother of existing?"

'underlayer'. As Locke (1690, Book II, Ch. 23, Section 2) pointed out, we haven't the faintest idea what substratum might be. This isn't surprising, given that substratum by definition isn't an idea at all, but the difference between a real object and the ideas (or properties) it 'contains'.

One of the few things we can say about substratum is that it is intimately connected with causation. A cause brings its effect into existence. We usually think that (with the possible exception of a 'first cause', an ultimate source of existence) an object cannot exist concretely without a cause. To ask for the cause of an abstract object, on the other hand, sounds very odd. One object seems therefore to acquire substratum from another object, and this 'spreading' of substratum through spacetime is what we call causation. Unfortunately this connection between substratum and causation does not shed much light on the notion of substratum, since causation is equally mysterious.

While physicalism does not involve an explicit rejection of substratum, its assumption of transparency amounts to such a rejection, for it thereby treats any non-transparent component of an object as trivial or inconsequential. Moreover, this rejection of substratum is rather counter-intuitive in its own terms. It leaves us without the resources needed to maintain a distinction (rather important to the common person, as well as the philosopher) between things that exist and things that do not.

A comparison between physicalism and Berkeleyan idealism is also instructive here. Berkeley's notion that ordinary material objects are just ideas strikes us as far fetched, yet it isn't nearly as silly as the view that *we persons* are just ideas in the mind of God. Berkeley himself ruled this out for reasons that are surely compelling. For, while ideas are components of conscious thoughts, it seems crazy to say that ideas *themselves* are conscious, as if fictional characters could have real experiences of their own. In a similar way, ideas themselves surely do not have intentionality or free will. As Berkeley said, ideas are passive and inert, not active like minds. According to physicalism, however, a person is an entirely intelligible object, corresponding exactly to mathematical ideas. Hence physicalism is contrary to these powerful Berkeleyan intuitions.

9. Conclusion

Physicalism and the mechanical philosophy share a common assumption that the world is fully transparent to a mind of sufficient power. Such transparency has long been judged impossible in the case of perceiving subjects, possessing consciousness and intentionality, though to my knowledge no one has previously demonstrated this impossibility.

I showed in Section 5 that a system with transparent physical states cannot have thoughts, as long as the semantic content of a thought in a given physical state is logically determined by that state. Such a supposition leads quickly to contradiction. In Section 8 I also argued that transparency is a dubious assumption in any case, for independent reasons.

References

Boyle, R. (1674) *The Excellency and Grounds of the Corpuscular or Mechanical Philosophy*.

Chalmers, D. (1996) *The Conscious Mind*, Oxford.

Hawking, S. (1988) *A Brief History of Time*, Bantam.

Jackson, F. (1982) "Epiphenomenal Qualia", *Phil. Quarterly*, vol. 32, no. 127, 127-136.

Leibniz, G. W. (1714) *Monadology*.

Lucas, J.R. (1961) "Minds, Machines, and Gödel", *Philosophy*, vol. 36, 112-127.

Locke, J. (1690) *Essay Concerning Human Understanding*.

Lockwood, M. (1991) *Mind, Brain and the Quantum*, Oxford: Blackwell.

Pascal, B. (1669) *Pensées*, Paris: Hachette, 1950, p. 44 (no. 72).

Stoljar, D. (2001) "Two Conceptions of the Physical", *Philosophy and Phenomenological Research*, 62: 253–281.

Strawson, G. (2008) *Real Materialism and Other Essays*, Oxford.