### CAN THE SELF BE A BRAIN?

#### ALAN KENNETH SCHWERIN

**Abstract.** Philosophical materialists suggest that a person can be identified with their brain. My paper is a critical investigation of this provocative thesis and an analysis of some of the prominent arguments to support this view. My overall argument is that there is more to this issue than some philosophers appear to acknowledge.

Keywords: Self; brain; materialism; grammatical field; Nagel; Parfit.

# Introduction: Language and the self

Who or what am I? When Renè Descartes proclaimed in 1642 that he is a self — he prefers to refer to himself as a mind or soul — he gave voice to a world-view or Weltanschauung that continues to dominate philosophical and religious circles:

For the present I am admitting only what is necessarily true: so "I am" precisely taken refers only to a conscious being; that is a mind, a soul (*animus*), an intellect, a reason — words whose meaning I did not previously know. I am a real being, and really exist: but what sort of being? As I said, a conscious being (*cogitans*). (Descartes 1969)

Descartes' account of a person — that has been derided as "the ghost in the machine myth" — has drawn the ire of many philosophers. (Ryle 2000) Ryle is certainly not alone in this respect. Most prominent among these nay-sayers are materialists such as Thomas Nagel, Patricia Churchland, J.J.C. Smart and David Armstrong. While their views, and the arguments that support their heretical positions vary, arguably the boldest case to be made for the materialists' cause is that that has been articulated by Thomas Nagel. In an unpublished rough draft that has been referred to and extensively analyzed by Derek Parfit in his *Reasons and Persons*, Nagel bravely suggests a striking response to the question, Who or what am I? (Parfit 1984) As Nagel sees it, he is his brain. Period. Or to be a little more precise, Nagel suggests that he is his brain in certain states.

We have much to gain from an investigation of the Nagel/Partfit discussion, even though it is based on an analysis of an unpolished account of Nagel's views. The arguments that inform Nagel's provocative thesis are some of the clearest and most forthright arguments by a materialist in the literature. In this paper I shall explore Nagel's view and its justification. Most importantly, I shall critically explore some of the fundamental issues that are raised by Nagel's position. To begin with, I shall

Principia 19(2): 235-246 (2015).

Published by NEL — Epistemology and Logic Research Group, Federal University of Santa Catarina (UFSC), Brazil.

discuss what I view as the most prominent of Nagel's arguments for his thesis, and then move on to consider the critical response to his thesis that Parfit articulates in his *Reasons and Persons*. Finally, I shall suggest that there is more at stake with Nagel's thesis than either Parfit or Nagel have explicitly acknowledged. In particular, I shall argue that Nagel's materialist thesis raises some fundamental questions on the relationships between the predicates of our language — issues, so it seems, that neither Nagel nor Parfit have accounted for in their respective views on the mind. Whether or not other materialists are prone to the same oversight, as interesting and important an issue it might be, is not of concern to me here.

But first a preliminary issue needs to be addressed. The philosophical problem on the relationship between the mind or self and the brain is enmeshed with a cluster of diverse fundamental issues. So any account of this relationship is not, and cannot be viewed as an isolated contribution to some self-contained philosophical problem. Parfit's *Reasons and Persons* views on the problem, and those of his antagonist, Nagel, prove no exception to this rule. As is intimated above, and will emerge more clearly shortly, Parfit and Nagel tacitly acknowledge this feature of the interrelationship between philosophical problems in their debate on the mind. For both garner arguments for their positions on the mind by appealing to their observations on the behavior of expressions one can use to refer to the brain and the mind. That is to say, by relying on their various *linguistic* investigations in their attempts to resolve what appears to be an *ontological* issue, both Parfit and Nagel are tacitly acknowledging the interconnectedness of issues that stem from essentially different realms.

This is not the place to investigate such fundamental moves by Parfit and Nagel. However, a more modest task is within our reach. What is in order here, in my view, is a limited assessment of their linguistic investigations into the mind/brain problem. As I hope to show in this paper, there is at least prima facie evidence to suggest that these investigators have overlooked certain important details of the linguistic expressions they are interested in. I shall argue that both Nagel and Parfit appear to have overlooked a fundamental feature of the relationship between predicates — a shortcoming that weakens, if not threatens to undermine both their current arguments on the mind. So my suggestion is that Nagel and Parfit — and by implication, others interested in the mind/brain problem — need to incorporate an account of this hitherto unnoticed feature of the relationship between linguistic expressions in their (linguistic) responses to the mind/ brain problem. This feature of linguistic expressions is illustrated and discussed in the third section of my paper. However, before we consider this (allegedly overlooked) feature, let me present the arguments of the two disputants — arguments that concern the referent of the word "I". To begin with, take Nagel's position.

# 1. One of Nagel's arguments for the thesis that he is his brain

What is the referent of the word "I"? Nagel has suggested that his brain is the referent of "I" when he uses the word. According to him, he may intend to refer to himself, the subject of particular experiences, but that when it comes to using "I", he is actually referring to his brain, irrespective of his intentions. So Nagel's claim can be spelled out as follows, as Parfit points out:

when I use the word 'I', intending to refer to myself, the subject of my experiences, *I am in fact referring to my brain*. (Parfit 1984, p.470, my emphasis)

But why does "I" refer to my brain, irrespective of my intentions to refer to myself? Or more importantly, what are Nagel's arguments for this claim?

According to Parfit, Nagel provides us with three arguments for his view: the first trades on the notion of essential properties and our ability to use language to refer to objects, the second leans on the claim that the referent of "I" is whatever explains my psychological continuity, and the third argument draws on our (likely) response to an imaginary situation where the survival of the brain does appear to be important. In this paper, I shall concentrate on the first of the three arguments mentioned here because it strikes me as the most important of Nagel's arguments from his point of view and because this is the argument Parfit spends most time on in his *Reasons and Persons*.

How then does Nagel's "essential argument" go? As I understand it, it works as follows. Take any object you wish that we are able to refer to. This object must possess some essential properties, for without them, that object could not exist as that object. And it's by virtue of these essential properties that we are able to use terms from our language to refer to that object. Hence the ability to refer successfully to any object presupposes that that object has the requisite essential properties. Take gold, for instance. Science has taught us that this element, and only this element, has the atomic number 79. So one, of perhaps a few, of the properties unique to gold is the atomic number 79. This then is one of its essential properties: every element with this atomic number must be gold, while elements without it cannot be gold. So when we use the term "gold" in any statement, as in "The gold chain on the table is expensive", the denoting phrase "the gold chain on the table" *must* refer to an item on the table composed of an element that has (at least) the atomic number 79, if that statement is to be true.

Here we must draw an important distinction between sentences and statements. Roughly, we can characterize the difference as follows: a sentence is a collection of (inanimate) signs or words arranged according to certain syntactic rules, while a statement is a particular *speech act* in which a sentence, usually a declarative sentence, is used to make an assertion that is either true or false. And as our ability to

make statements or assertions presupposes our ability to refer to, or to describe, or predicate various phenomena, events or entities in our universe, the successful use of a (probably declarative) sentence rests on the successful performance of the speech acts of referring and predicating. As John Searle, in his book *Speech Acts* puts it, in making a statement one is performing a particular illocutionary act that entails that some propositional act of referring or describing has occurred. As he points out,

the characteristic grammatical form of the propositional acts are parts of sentences: grammatical predicates for the act of predication, and proper names, pronouns and certain other sorts of noun phrases for reference. (J. Searle, *Speech Acts*, p.25)

It is necessary to go one step further here in this short digression, for statements are used not only to perform illocutionary acts, but also for the performance of various perlocutionary acts, *i.e.* speech acts that are intentionally designed to elicit certain responses, behavioural or otherwise, from one's audience. Thus when a speech act has an effect on the actions, thoughts, beliefs or images of another, these consequences can be viewed as perlocutionary effects associated with the speech act in question.

However, it is clear that not all statements, that necessarily are illocutionary acts, are accompanied by perlocutionary effects. That is to say, not all illocutionary acts are perlocutionary acts as well. For instance, when a statement occurs in a stipulative definition, one is not engaging in a perlocutionary speech act. Thus when I introduce you to the referent of "pen", as in the statement "This is a pen", no perlocutionary effects are intended. Or when the chemist informs his class by stating "Salt contains the chemical sodium", he is performing an illocutionary act that has no (intentional) perlocutionary effects. On the other hand, when the scientist states "Electrons orbit through various electromagnetic fields", not only is he performing an illocutionary act when making his statement, his statement also conjures up certain ideas or images of fields within which certain items, namely electrons, move. So, as with statements like this one, a perlocutionary effect can accompany the illocutionary act of making a statement.

In the discussion that follows I shall observe the distinctions introduced here. For the sake of clarity, from now on let us say that a statement that is intentionally free of any perlocutionary effects is a *scientific statement*, or simply a statement. Should there be occasion to talk about the second type of statement, that is to say, where any perlocutionary effects are intended by the speaker to accompany a statement, we shall speak of *metaphorical scientific statements*, or simply metaphorical statements.

To return to Nagel's argument. We have seen that for him, a statement, that is to say a scientific statement, *must* have a denoting phrase or singular term that refers to an item in the world if that statement is to be true. Now the term "I" is used in

true statements. For instance, the statement "I am happy today" is true as I type this. So the term "I" must successfully refer to at least one, of perhaps a few essential properties of some item in the universe. All of which raises two important questions:

Q1: What are the essential properties being referred to by the term "I"?

Q2: Which item in the universe has these essential properties?

In answer to the first question, Nagel suggests his mental states are essential to him. And which item of the universe has these mental states? Nagel's response is initially noncommital: "whatever in fact makes it possible for the person TN to identify and reidentify himself and his mental states". (Parfit 1985, p.469). So which item satisfies this requirement? According to Nagel it is his brain, for without it, he in fact will not be able to identify and reidentify himself. As he puts it, if certain (physiological) states and activities of his brain underlie his mental capacity, "then that brain in those states ... is what I am, and my survival of the destruction of my brain is not conceivable". (Parfit 1985, p.469). In the light of all of this, Nagel concludes that his use of the word "I" entails that the word refers to his brain i.e. that he is his brain in particular states.

The details of Nagel's argument could be highlighted more graphically, and ought to receive more attention than I have given it. However, we need not be concerned with this task, for the exposition we have here of his views is adequate for the purposes of this paper. We now need to consider — in broad outline — Parfit's views on the matter. As I hope to demonstrate later, *both* Parfit and Nagel, who hold opposing views on the question whether I am my brain, have overlooked a fundamental assumption: the assumption that it makes sense to claim that I am my brain. So before we consider this important issue and some of the related concerns that the implicit commitment to this view raises, let's see what Parfit has to say about the matter.

# 2. Parfit's response to Nagel

Parfit does not agree with Nagel that he is his brain. According to Parfit, he is not his brain. As he says, a

person is an entity that has a brain and body, and has different experiences. My use of the word "I" refers to myself, a particular person, or subject of experiences. *And I am not my brain*. (Parfit 1985, p.471, last emphasis mine)

So not only does he deny that he is his brain, Parfit also asserts that when he uses the word "I", the referent is not some physical entity, such as a brain, but himself — a subject of experiences. On what grounds does Parfit make these assertions?

As we saw earlier, Nagel suggests that the actual referent of "I" when he uses the word, is his brain, irrespective of his intentions to refer to himself, the subject

of his experiences. In his response, Parfit's primary objective is to demonstrate that this premiss of Nagel's argument need not be true. Parfit wants to show that "we *can* refer to ourselves". (Parfit 1985, p.473). In short, Parfit's claim is that references to our brains are not necessary when we use the word "I". His argument against Nagel works as follows.

Consider the following situation. Jack walks into the room, and after a careful inspection of the furniture, pointing to the only table in the room, utters the following statements:

- S1: The table in that corner is a poor conductor of electricity.
- S2: The table in that corner is beautiful.

Suppose now that both statements are true. What then are the referents of the denoting phrases that occur in these statements? Take the first statement. This statement is clearly about the properties of the various elements that collectively constitute the table. The table is composed of constituents (atoms) that are poor conductors of electricity. So the denoting phrase of S1, namely "the table in that corner", strictly speaking, refers to the constituents of the table. Now consider the second statement. Statement S2 is about the table itself, and not, for instance, about the constituents of the table. For it is the table, and not its atoms that is held to be beautiful. Suppose Jack's statement was prompted by the fact that the table has been finely carved. When he asserts statement S2 he is not saying that the atoms of the object in that corner are beautifully carved. No. What statement S2 does, is tell us that Jack has found an aspect of the object in that corner attractive, namely its aesthetic appearance, and that he wants to draw our attention to it. So statement S2 is an instance where its denoting phrase, namely "the table in that corner", does have a referent, but the referent is the (aesthetically satisfying) table. and not, for instance, the constituents of the table, as was the case with the denoting phrase of S1.

This is interesting. For this is an instance where two statements are about *different* phenomena, in that the referents of the denoting phrases are different, even though the *same* denoting phrase occurs in the two statements. As statements S1 and S2 illustrate, a denoting phrase on one occasion may refer to the constituents of an entity, while on another occasion the same denoting phrase can be used to refer to the entity itself, and not to its constituents. The moral to be drawn from this, or rather two of them are the following;

- (i) The denoting phrases of true statements are not referentially determinate: their referents vary from context to context.
- (ii) The denoting phrases of true statements do not carry their referents on their sleeves: for we cannot tell from the denoting phrase alone what its referent is.

But denoting phrases are only one, of possibly many types of linguistic expressions used to refer to objects. Can the results here be extended to these other linguistic forms used to refer to entities? In particular, do the results here have any bearing on the word "I"?

For Parfit they do. For if the word "I" is an expression that has a referential function, as do proper names and denoting phrases, for instance, it is surely reasonable to view the outcome of the analysis on these other referring expressions as applicable to the word "I"? Parfit clearly assumes that this is a reasonable course of action to take, for he takes it! As he says, "we can make similar claims about our use of the word 'I". (Parfit 1985, p.472) But if the word "I", as with denoting phrases and proper names, is referentially indeterminate, the brain need not be its referent. "I" may well have other referents as well. So "I", contrary to Nagel's view, might after all refer to myself, the subject of my experiences. From this Parfit concludes that we "can retain our natural belief that we can refer to ourselves". (Parfit 1985, p.473)

It is important to be clear about what has, ostensibly, been achieved here by Parfit. What he has attempted to do, is establish the possibility that "I" does not refer to the brain. Or as I put it earlier, Parfit wants to provide reasons for suggesting that Nagel's claim need not be true. So Parfit has not tried to show that Nagel's view is false:

I have tried to answer Nagel's arguments. *My answers do not show that his view is false*. But I believe that they show that we can, defensibly, take a different view. The question remains open. (Parfit 1985, p.477, my emphasis)

Thus Parfit is suggesting that though he has not shown that Nagel's view is false, someone else might settle the matter one way or the other. But this suggestion rests on the fundamental assumption that Nagel's view makes sense in the first place. For unless his view makes sense, attempts to either confirm or falsify Nagel's claim are otiose. So the question arises: does the claim that I am my brain make sense?

#### 3. One condition sensible statements must meet

When does a statement make sense? Does the statement "I am my brain", make sense? And most importantly, *can* the statement "I am my brain" make sense? These are certainly demanding questions, and warrant a good deal of attention — far more than I can offer here. Nevertheless, I think that some inroad into these issues can be made in this paper. For as I hope to demonstrate below, it is possible to isolate one of the conditions that a statement must meet for it to qualify as a sensible statement. Equipped with this necessary condition, it remains for us to hold this condition up against Nagel's thesis to determine whether or not it makes sense.

Consider the following two statements, one of which makes sense. Assume that the statements here are being used as scientific statements, and not as metaphorical

statements. That is to say, in relying on the two sentences Sa and Sb to utter two statements, one is performing two illocutionary acts, without intending any perlocutionary effects. (On this point look back on my earlier discussion in section one on the distinction.)

Sa: Peter is a man.

Sb: London is my sister.

Why does the first statement make sense, while the second does not? As it happens, the words "Peter" and "London" both have straightforward referents: the word "Peter" refers to my brother-in-law, while the word "London" refers to a city situated alongside the river Thames in England. So in both statements the expressions occupying the subject position do have referents: Peter and London respectively. But statement Sa makes sense, while statement Sb does not. How do we account for the difference?

At least one reason for this discrepancy, I suggest, is that while the predicate term of the first statement (i.e. "a man") can sensibly be applied to the entity referred to by the word "Peter", the predicate term of the second statement (i.e. "my sister") cannot be applied in any sensible manner to the referent of the word "London". Cities are not entities of which it makes any sense to apply predicates like "my sister". In addition, if it does make sense to apply the predicate term "a man" to an entity, it also makes sense to apply *other* appropriate predicates to that entity as well. For instance, if the term "man" is applicable to Peter, the terms "human being", "intelligent", "honest" and "shrewd" can also be sensibly applied to Peter. Let me say a little more on this last point.

For clarity of exposition, I shall coin a phrase to describe this phenomenon of language that I am alluding to here. Let's say that terms belong to various *grammatical fields*, such that, given the assertion or denial of a statement that makes sense, if one understands that language, one is able to specify possible alternative, yet related, terms from the grammatical field that are applicable to the entity referred to by the statement asserted, or denied. This phenomenon of language is graphically illustrated when we consider negative statements. Take the following negative (scientific) statement:

Sc: That is not red.

If this statement does make sense, the denial that the referent of "that" is red, entails that the entity referred to by the statement is blue, green, yellow, *or some other colour*, and not, for instance, that the entity is a baseball, or three years old. (The statement "That is a baseball" does not sensibly follow from the statement "That is not red", for instance.) So, if it makes sense to deny that the term "red" is applicable

to the referent of "that" — whatever it may be — at least some other term from the grammatical field occupied by the word "red" can be sensibly applied to that referent. From this it appears to follow that if some other term from the grammatical field occupied by "red" cannot be sensibly applied to the referent of a statement, the term "red" itself cannot sensibly apply to the referent either. For instance, with the statement

## *Sd*: The sound of his voice is green,

the term "green" cannot be sensibly applied to the entity referred to by the denoting phrase "the sound of his voice". But the term "green" occupies the same grammatical field as that of "red". Hence the term "red" cannot sensibly apply to the referent of statement Sd.

Now if the term "green" cannot be sensibly applied to the referent of Sd, this statement makes no sense. Which suggests that any other statement that relied on a term from the grammatical field occupied by "green", to apply to the same referent as that of statement Sd, would also fail to make sense. Thus the statement,

### Se: The sound of his voice is yellow,

also fails to make sense. Statement Se fares no better than statement Sd, where sense is concerned.

In the light of all this, I think we can articulate one of the conditions any statement with a predicate term must meet for it to qualify as a sensible statement. We have seen that if a predicate term cannot sensibly be applied to the referent of a statement, that statement does not make sense. Secondly, I have suggested that predicate terms occupy grammatical fields, such that if a term can sensibly be applied to an entity, any of the other terms from its grammatical field must also be sensibly applicable to that entity. This suggests the following:

**Condition1:** For any two statements S1 and S2, with predicate terms t1 and t2 respectively, if statements S1 and S2 are about the same referent r1, and terms t1 and t2 occupy the same grammatical field F1, if either term t1 or term t2 cannot be sensibly applied to referent t1, neither statement t1 nor statement t1 make sense.

My analysis above, and especially the articulation of Condition 1, can prove very useful here, I think. For it now places us in the position to see what Nagel's thesis amounts to.

If my analysis is correct, we realize that Nagel, in asserting his thesis "I am my brain", is (tacitly) asserting that the term "my brain" is applicable to the referent of this statement — whatever it may be. But this is to presuppose that the term "my

brain" can be sensibly applied to this referent. Or to put it another way. Nagel is assuming that it is true that the predicate term "my brain" is satisfied by the referent of the statement "I am my brain". Given this assumption, it follows from Condition 1 that other statements that have the same referent, but that use different terms from the grammatical field occupied by "my brain", *must* also be statements that make sense i.e. they must be scientific statements that contain predicate terms that can be sensibly applied to the referent of "my brain". All of which gives rise to a crucial question: can the other terms from the grammatical field occupied by "my brain" be sensibly applied to the referent of Nagel's statement? If not, according to Condition 1, it would appear that the statement "I am my brain" cannot make sense. And if Nagel's thesis cannot make sense, it cannot be verified. Without further ado, let us consider this important issue.

# 4. On the sense of Nagel's thesis

Nagel, in asserting the statement "I am my brain", is suggesting that it does make sense to apply the term "my brain" to the referent of the word "I". But does it? One way to help decide the issue, as has been suggested in the previous section of this paper, is to identify the grammatical field occupied by the term "my brain", and to see whether or not the terms from this grammatical field are sensibly applicable to this referent. And to do this, given Condition 1, we can consider *any* of the terms from the grammatical field occupied by "my brain". Following this procedure, I think that one can articulate an argument to show that Nagel's claim does not make sense. The argument proceeds as follows.

Assume that Nagel's claim does make sense. Now suppose that some future investigator establishes that this claim is false. (Remember, Parfit pointed out that the matter was an open one, presumably to be settled one way or the other at some future date.) In that case the following statement is true:

## Sf: Nagel is not a brain.

But if it does make sense to deny that he is a brain, it follows that the assertion of some other statement, that relies on other terms from the grammatical field occupied by the term "my brain", will at least also make sense. (Of course, this alternative sensible statement may even be true.) So what are some of the other terms in the grammatical field occupied by the term "my brain"? The term "my brain" occupies the same grammatical field as that occupied by "nose", "mouth", "lung", "heart", "left arm" and other (human) anatomical terms. For instance, in an anatomical class the instructor's statement,

Sg: That is not a lung, but a heart,

made while pointing to an object on the table, clearly can make sense, because "lung" and "heart" occupy the same grammatical field. Now, if the statement "Nagel is not a brain" does make sense, the following must also make sense,

### Sh: Nagel is a left arm.

For if "brain" and "left arm" occupy the same grammatical field, as it seems they do, statements with the same referent that used these predicates must all make sense if one does. But statement *Sh* does not make sense. For Nagel is a *person* — an entity that is capable of loving, dreaming, forming intentions, and thinking of others, for instance. And it certainly makes no sense to suggest that a left arm could love, dream, form intentions and think of others. So while it makes sense to claim that Nagel *has* a left arm, it does not make sense — or is false — to claim that the term "left arm" can be instantiated by a person.

But if statement Sh fails to make sense, it seems that statement Sf does not make sense either, according to Condition1: for the terms "brain" and "left arm" occupy the same grammatical field. But statement Sf and Nagel's claim both have the same referent i.e. Nagel, and rely on the same predicate term "my brain". So if statement Sf does not make sense, it seems that the same must be said for Nagel's thesis. From this it appears to follow that Nagel's claim, "I am my brain", does not make sense.

Is this a convincing argument? More particularly, is it correct to conclude from this that Nagel's thesis *cannot* make sense? Probably not, for at least the following two reasons.

- a) On the one hand, the argument above rests on a specific contingent suggestion concerning a grammatical field occupied by the word "brain". Further empirical research into the issue may reveal that the members of some linguistic communities operate with grammatical fields different to that relied on by the proponent of the argument.
- b) On the other hand, the specification of the current status of a grammatical field is not necessarily a specification of its future status. For all we know, the thesis "I am my brain" is the harbinger of a reorientation of the grammatical fields occupied by "brain" and "mind".

All this suggests that it remains to be seen whether or not the above contingent suggestion on the grammatical fields occupied by "brain" and "mind" is as widely accepted, and eternal, as the stronger (negative) assessment of Nagel's claim would require. So scientific statements on the relationship between the self and the brain must remain indeterminate, if not meaningless. From this it follows that the proposal that the self can be a brain must remain an open question.

## Acknowledgments

I must thank Derek Parfit, Mark Kulstad and Dagfinn Føllesdal for their constructive responses to earlier versions of my paper.

#### References

Armstrong, D. 1968. *A Materialist Theory of the Mind*. London: Routledge and Kegan Paul. Churchland, P. 1989. *Neorophilosophy: Toward a Unified Science of the Mind-Brain*. Cambridge: The MIT Press.

Descartes, R. 1969. *Descartes: Philosophical Writings*. Translated by Elizabeth Anscombe and Peter Geach. London: Thomas Nelson and Sons.

Nagel, T. 1986. The View from Nowhere. New York: Oxford University Press.

Parfit, D. 1984. Reasons and Persons. Oxford: Oxford University Press.

Ryle, G. 2000. The Concept of Mind. Chicago: University of Chicago Press.

Searle, J 1969. Speech Acts: An Essay in the Philosophy of Language. Cambridge University Press

Smart, J. J. C. 1963. Philosophy and Scientific Realism. London: Routledge and Kegan Paul.

ALAN KENNETH SCHWERIN
Associate Professor of Philosophy
Monmouth University
New Jersey
USA
aschweri@monmouth.edu