NATHAN STEMMER

THE HYPOTHESIS OF OTHER MINDS: IS IT THE BEST EXPLANATION?

(Received 20 December, 1985)

I

In a recent paper, Pargetter (1984) discusses the justification of the hypothesis that other people have mental lives. He presents various convincing arguments which show that basing this hypothesis on the analogy between myself and other people is of little value. First, this analogical inference is based on a single instance. Second, the conclusion of the inference is uncheckable. Finally, analogical inferences can only reasonably be used in cases where there are sufficient relevant similarities and no relevant differences between the cases where the projected property is known to hold and the cases where that property is being inferred to hold. Now there are many similarities between me and other people; but there are also many differences. While we do believe that these differences are not really very important, "it is very hard to see how, without begging the question, we can argue that these differences are not relevant as far as an analogical inference is concerned" (p. 160). One difference is particularly striking, namely, that "the evidential base is about *my* case, *my* instances of pain accompanying pain behavior, etc., while the conclusion is about cases which are *not mine*" (p. 160).

But Pargetter believes that we can still justify the hypothesis of other minds by treating it as the conclusion of an inference to the best explanation rather than of an inference by analogy. For suppose we see a man with a deep cut in his hand. The cut is bleeding. The man is clutching the cut hand with his other hand. He looks pale and tense, grunts and groans, and utters sentences such as 'My hand is hurting' and 'I am in pain'. In this case, Pargetter believes that the best explanation of the man's behavior is the hypothesis that the man is in pain (where being in pain is supposed to be a mental state), and his pain is much the same as our pain in similar circumstances.

In this paper, I shall argue that Pargetter's conclusion is unwarranted.

Philosophical Studies **51** (1987) 109–121.

Although the hypothesis that other people have minds does explain our own mental feelings as well as the behavior of ourselves and of other people, it is unlikely that this is the best explanation of these phenomena. There exists an alternative hypothesis which explains these phenomena without assuming the existence of other minds. Yet, there are strong reasons which suggest that the alternative hypothesis gives a better explanation of the phenomena than the hypothesis of other minds.[1]

II

An hypothesis must satisfy various criteria in order to be the best explanation of a class of phenomena. Pargetter admits that we may not be able to lay down completely adequate criteria for selecting the best explanation from a set of alternative explanations. But he rightly states that in many cases there is general agreement when such choices need to be made (p. 159). He later mentions one of these criteria, namely, that the explanation should take into account "the total available evidence" (p. 162). Since most scholars agree with Pargetter that this is indeed an important criterion for the evaluation of explanations, let us therefore examine Pargetter's hypothesis of the existence of other minds in the light of this criterion.

Suppose that some days later we see that the man who had cut one of his hands now cuts one of his legs. He again grunts and groans, but instead of clutching his cut hand with his other hand he now clutches his leg with this hand. In this case, it is not sufficient for Pargetter to assume that the man is now in *some* mental state. He will also have to assign a very high degree of probability to the hypothesis that this state is different from the previous state. And if we later observe the man eating a cookie and showing his pleasure with some appropriate behavior, then, again, Pargetter will have to assume not only that the man is now in some mental state but that this state is likely to be different from the other states. Making such distinctions between mental states is necessary if Pargetter wants his explanation to take into account the *total* available evidence. For this evidence tells us not only that, under certain conditions, people show *some* behavior suggesting that they are in a mental state, but that the behavior is very often significantly different. This shows that if Pargetter wants the hypothesis of other minds to account for all the available evidence, he must try to make appropriate distinctions between different mental states.

Now, Pargetter has probably no objections to this conclusion. It seems to be implicitly and even explicitly accepted by the philosophers who assume that people have minds. These philosophers, whom I shall call *mentalists*, speak of headaches and stomachaches and they distinguish not only between these two types of pains, but also between them and pleasures such as eating cookies or listening to a concert, and they again distinguish these mental states from, e.g., mental belief states such as when a person believes that ravens are black.

This suggests that the following is a fair account of the mentalists' explanation of the behavior of the man who hurt his hand and his leg. By observing that he was hurt in different places they temporarily advance the hypothesis that the experiences produce a particular mental state *a* in the first case and a different mental state *b* in the second case. If it is then observed that in the first case the man clutches his hurt hand, while in the second case he clutches his hurt leg, then mentalists will greatly increase the degree of probability which they assign to the hypothesis that *a* and *b* are indeed different mental states.

We notice that mentalists must always base their hypotheses regarding differences between the mental states of other people on indirect evidence, since these states are not directly observable – in Pargetter's terms, they are uncheckable. The indirect evidence will generally consist of differences between the experiences or events that give origin to the mental states and of differences between the behavior that is caused, or accompanied, by these states.

The evidence will be indirect even if a person directly tells us that he is in a particular mental state and even if we believe that he is sincere. Suppose the person utters 'I am in pain'. Mentalists will then assume that the person refers with the word 'pain' to the mental state pain. But if we consider that the person had to *learn* that the (English) word 'pain' refers to this mental state, then we realize that the utterance of 'I am in pain' gives us only an indirect instrument for individuating the state. For suppose the person has learned to apply the word 'pain' to a mental state *m*, by hearing in his childhood his parents saying 'pain' while they were assuming he was in pain. Since his parents could only use indirect evidence for determining whether he was really in pain when they said 'pain', it is of course possible that the state *m* is different from the state pain. And the evidence given by 'I am in pain' will also be indirect if the person has learned to apply 'pain' to state *m*, by hearing

utterances of the word when it was being applied to other people. For in this case, he could observe only the behavior of the other people, which he may have interpreted incorrectly. Therefore, even the verbal reports of a person's own mental states give only an indirect individuation to the state which mentalists attribute in this case to the person.[2]

With respect to the evidence we have discussed so far, the *detailed* version of the hypothesis of other minds – i.e. the version which intends to distinguish between different mental states of other people – seems to be a good, and perhaps even the best explanation. But we shall now see that this changes radically once we take into account additional available evidence. This is the evidence which shows a great number of significant correlations between certain processes in the neurophysiological systems of human beings and the behavior on which mentalists rely for postulating mental states in these people. It has been observed, for example, that severing certain neural pathways has often produced a decrease in the groaning and grunting of a person who has cut his hand, or that certain types of brain damage have caused changes in the number of a person's belief states, as evinced by relevant behavior, including verbal behavior.

How do mentalists account for these evidential data? The usual approach, and I think that Pargetter would agree, is to assume that in addition to the mental states, there also exist neurological states (or processes) that play a direct or indirect role in producing the relevant behavior. Different alternatives arise within this approach. According to one of them, certain experiences or events – e.g. cutting one's hand or eating cookies – produce particular neurological (or neurophysiological) states in the person. These states give origin to certain mental states which finally produce the relevant behavior, perhaps together with the neurological states. Another alternative is to assume that the neurological states directly cause the behavior, and the mental states are only perceived by the person, but play no causal role in producing the behavior. Or one might assume that the experiences or events directly produce the mental states. These mental states give origin to the neurological states which then produce the behavior, perhaps together with the mental states.

Usually, the neurological states will not be directly observable. One must therefore use indirect evidence for distinguishing between different neurological states. That is to say, most, if not all of these states will receive only an indirect individuation.

I have mentioned different procedures for expanding the detailed hypothesis of other minds in order to account for the neurophysiological data. It does not matter whether each of these alternatives has actually been proposed by mentalists.[3] What is important for us is that if we want the hypothesis of other minds to be the best explanation of the relevant phenomena, it must account for the neurophysiological data – they are part of the *total available evidence*. We must therefore expand the hypothesis by attributing to people neurological states (or processes), in addition to mental states. Let us call the hypothesis, which is obtained by adding an appropriate neurological (or neurophysiological) component to the detailed hypothesis of other minds, the *mentalist* hypothesis.

But once the hypothesis of other minds has become the comprehensive mentalist hypothesis – and only in this form can it be a candidate for being the best explanation of the relevant phenomena – we can obtain an alternative hypothesis which explains the phenomena in a better way. This alternative hypothesis will be called the *neurophysiological* hypothesis, in short, the *neurological* hypothesis.

III

The neurological hypothesis is obtained in the following way. We eliminate from the mentalist hypothesis all references to the mental states for which the hypothesis has corresponding neurological states (or processes).[4] We then replace the names of the remaining mental states $m_1, ..., m_k$ $(0 \leqslant k)$ by an equal number of new names which are supposed to refer to the indirectly observable neurological states $n_1, ..., n_k$.[5] We now use all the data which mentalists use for individuating the mental states of the mentalist hypothesis, in order to individuate with the help of these data the corresponding neurological states of the new hypothesis. Finally, we also use for the individuation of the neurological states all those additional data that may be provided by neurophysiology.

To illustrate the neurological hypothesis, let me give a sketch of how we would explain what occurs when a person cuts his hand. Instead of assuming that (according to one version of the mentalist hypothesis) this experience produces the neurological state d in the person which then produces the mental state a, we now assume that the experience only produces the neurological state d. And instead of assuming (according to one version of the

mentalist hypothesis), that the mental state *a* is the cause of the relevant posterior behavior, we now assume that this behavior is caused by the neurological state *d*. Moreover, we individuate the neurological state *d* by using the same data which mentalists use for individuating the mental state *a*. Finally, we use the data which are provided by neurophysiology (if any) in order to increase the precision of our individuation of the neurological state *d*.[6]

It is not difficult to see that this neurological explanation is simpler than the one which is based on any version of the mentalist hypothesis. Instead of assuming the existence of two indirectly observable entities – the neurological state *d* and the mental state *a* – we now assume only the existence of one such entity – the state *d*. And instead of having to give an individuation to the two states *d* and *a* by relating them to certain observable events or behavior, we now have to individuate only state *d*. Moreover, the neurological explanation has at least the same degree of precision as the mentalist explanation, since it makes use of all the data which mentalists use for individuating the mental state *a*, in order to individuate the neurological state *d*. And if there exist additional neurophysiological data that can be used for this purpose, then we can even increase the degree of precision of the neurological explanation.[7]

IV

I shall now argue that the neurological hypothesis gives a better explanation to the available evidence than the mentalist hypothesis, i.e., the detailed and expanded version of Pargetter's hypothesis of other minds. The argument will be based on widely accepted methodological criteria.

As said earlier, we may not be able to state precise criteria for selecting the best explanation from a set of alternative explanations for a particular corpus of data. But there is general agreement, and I believe this does not exclude Pargetter, that degree of simplicity is an important criterion. Now, the (detailed and expanded) mentalist hypothesis assumes the existence of two kinds of indirectly observable entities: mental as well as neurological states. But the neurological hypothesis only assumes the existence of one kind: the neurological states. And although the number of the indirectly observable entities which are assumed by the mentalist hypothesis is perhaps not twice the number of those assumed by the neurological hypothesis,

it is clearly higher. The mentalist hypothesis must therefore specify more connections between indirectly observable entities and observable entities than the neurological hypothesis. It follows that the former hypothesis has a higher degree of complexity than the latter. It assumes more indirectly observable entities, and it must specify more connections between these entities and observable entities. On the other hand, the observational content of the mentalist hypothesis is not greater than that of the neurological hypothesis. All the observational data that are explained and predicted by the former hypothesis are also explained and predicted by the latter. Consequently, the greater complexity of the mentalist hypothesis, which derives from its large mentalist apparatus, is not compensated by an increase in explanatory and predictive power of observational data. By applying widely accepted criteria for theory evaluation, we can therefore conclude that the neurological hypothesis gives a better explanation to the available evidence than the mentalist hypothesis. The former is simpler; yet, its observational content is the same, and perhaps even larger.

Our analysis shows that Pargetter's conclusion regarding the hypothesis of other minds being the best explanation is unjustified. Once we take into account the *total* available evidence – and Pargetter agrees that this is the correct approach for receiving good theories – it becomes clear that the neurological hypothesis gives a better explanation to this evidence than the hypothesis of other minds.

Still, we should not underestimate the importance of Pargetter's move. By giving the hypothesis of other minds the status of the conclusion of an inference to the best explanation, we can treat it in the same way as other scientific hypotheses. This allows us to apply the usual criteria of theory evaluation to the hypothesis. And since the application of these criteria in other fields of scientific inquiry has very often enabled us to make real progress in these fields, by adopting Pargetter's approach we increase the probability of also making real progress in the present field. (The fact that the criteria of theory evaluation point to the superiority of the neurological over Pargetter's mentalist hypothesis is, of course, beside the point.)

The neurological hypothesis has many points in common with so-called eliminative materialism (see, e.g., Churchland, 1981). But it also differs from this position, for the hypothesis does not reject the useful scientific conclusions of mentalist 'folk psychology'. It only changes the *interpretation* of the mentalist conclusions. Instead of attributing certain effects to mental

entities or instead of assuming that certain events have mental consequences, the hypothesis attributes the effects to corresponding neurological entities or assumes that the events have corresponding neurological consequences. But all the lawful connections between observable entities or events that have been established by mentalists remain valid. Consequently, the neurological hypothesis – which corresponds to what one might call *interpretative eliminative materialism* – exploits all useful scientific results that mentalists have been able to obtain. In addition, of course, it also employs all useful neurophysiological results.

V

I shall now briefly examine three further arguments that have been given in support of the hypothesis of other minds. We shall see that these arguments do not reduce the explanatory superiority of the neurological hypothesis.

The first argument, which is also mentioned by Pargetter (p. 159), compares the mentalist hypothesis with other hypotheses which also postulate the existence of entities that are not directly observable, such as hypotheses about atomic particles. A great number of these hypotheses have been widely accepted by the scientific community. This shows the scientific legitimacy of postulating the existence of mental entities, even though these entities are not directly observable.

But besides being scientifically legitimate, an hypothesis must satisfy other conditions in order to be the *best* explanation of a class of phenomena. And here there is a fundamental difference between hypotheses about, e.g, atoms and electrons and the hypothesis of other minds. The former have at least one of the following characteristics: (a) they are simpler, or at least no more complex, than the known alternative hypotheses that cover the same phenomena, or (b) they make it possible to explain or predict more observable phenomena. This gives the hypotheses not only the status of being scientifically legitimate, but also of being good explanations. But we have seen that the mentalist hypothesis satisfies none of these conditions. It is much more complex than the neurological hypothesis; yet, this great increase in complexity has no observational consequences. The (observational) explanatory and predictive power of the former hypothesis is not greater than that of the latter hypothesis. We can therefore conclude that although the assumption of the indirectly perceptible mental states does not prevent the

mentalist hypothesis from being scientifically legitimate, it is not sufficient for making it the best explanation of the relevant phenomena. On the contrary, the discussion in the previous section shows that the neurological hypothesis gives a better explanation to these phenomena.[8]

The second argument in favor of the hypothesis of other minds is based on the fact that humans are apparently able to give quite reliable verbal reports about their mental states. For example, Price (1960) states that although the things one discovers by introspection are private ones, "they are nonetheless publicly describable, since the information one gets by means of introspection can be imparted to others" (p. 81). Now, if people can give us verbal information about their mental states, then these states indeed appear to exist. This would give strong support to the hypothesis of other minds.

In Section II, I have discussed some aspects regarding the reliability of such verbal reports. But I would now like to discuss a much more fundamental issue that is involved here.

It is clear that for the argument to support the hypothesis of other minds, we must assume that utterances of sentences such as 'I am in pain' give us information about the *mental* states of a speaker. But if we look at the process of language acquisition from a neurophysiological point of view,[9] then there is no reason to accept this assumption. Instead of attributing the utterance of 'I am in pain' to the fact that the speaker is in a particular mental state (which may correspond to a particular neurological state), neurophysiologist theories attribute the utterance to physiological and neurological causes, say, to the neurological state n which is produced by certain physiological factors. According to neurophysiologist theories, therefore, the utterance of 'I am in pain' gives us information about physiological and neurological events that affect the speaker, rather than about his mental states.

We have seen earlier that neurophysiological theories are simpler than mentalist theories. This also holds for theories of language acquisition (and language use). The neurophysiological theories of language acquisition, which are derived from mentalist theories according to the method described in Section III, assume the existence of less indirectly perceptible entities than the corresponding mentalist theories. Yet, they explain and predict the same observational phenomena, including the correlations between utterances of sentences such as 'I am in pain' and certain types of behavior. We can therefore conclude that neurophysiological theories of language acquisition are

better than mentalist theories. They give a simpler explanation to the available evidence.

Important regularities have been observed between utterances of sentences such as 'I am in pain' and certain types of behavior. If we accept a mentalist theory of language acquisition (and of language use), then these regularities indeed support the hypothesis of other minds, since in this case we conclude that the utterances of the sentences give us information about the mental states of a speaker. When accepting neurophysiological theories, however, this conclusion is no longer valid. We do acknowledge the important regularities that have been observed between the utterances and certain types of behavior. But we now conclude that these utterances give us information about neurophysiological events that affect the speaker, rather than about his mental states. Since neurophysiological theories give a simpler, and therefore a better, explanation of the relevant evidence than mentalist theories, we are justified in accepting their conclusions. We can therefore conclude that the regularities between the utterances and the behavior do not give special support to the hypothesis of other minds. Rather, they are part of the evidence which supports the parsimonious neurophysiological theories.

The third argument does not directly support the hypothesis of other minds. Rather, it suggests that the neurological hypothesis may fail to account for certain types of phenomena. The argument is best illustrated by an example given in Jackson's (1982) discussion of the perception of so-called qualia. Suppose that Mary has spent her whole life in a black and white room. With the help of a black and white TV monitor she has learned all that science, including neurophysiology, could ever convey about color-perception. Now suppose that Mary leaves the room for the first time and sees a clear blue sky. It seems obvious that, in spite of the fact that Mary has already learned everything about color-perception, she learns something new by seeing something blue for the first time. This suggests that there are certain phenomena which are not accounted for by the neurological hypothesis.

According to the neurological hypothesis, states of knowledge are neurological (or neurophysiological) states. Hence, the state which corresponds to Mary's knowing everything about color-reception, as well as the state which is produced by Mary's seeing the blue sky, are neurological states. Since at this stage of technical development, none of these states is directly observable, we have to give them an indirect individuation. To this effect, we shall

use all relevant data which we can obtain, including those used by mentalists for individuating perceptions of qualia. Now, whether these data will establish that the states are indeed different is difficult to say; it is an empirical issue. But if there are data which directly or indirectly confirm Jackson's claim that Mary has learned something new – e.g., she may now utter the sentence 'Seeing something blue is experienced thusly' (cf. Conee, 1985, p. 300) – then the neurological hypothesis will indeed support the conclusion that the states are different. We thus see that the neurological hypothesis can account for the perceptions of qualia, provided there are data which allow us, at least theoretically, to individuate the corresponding neurological states. (Clearly, the proponents of mentalist hypotheses also need such data, since they must have means for distinguishing between the perceptions of different kinds of qualia in order to give the hypotheses a good explanatory status.)

Let me close by stressing that the method of Section III, which transforms the mentalist hypothesis into the neurological hypothesis, is to be applied to all mental items, including my own feelings, thoughts, states of self-awareness, states of knowledge, expectations, desires, etc. (if I would wish to treat these items as mental entities). According to the neurological hypothesis, therefore, when I am in a pain state, I am merely in a particular neurological state. And since it is I who am in this state – I have privileged access –, it is I who am doing the grunting and groaning, or the uttering of sentences such as 'I am in pain' (if I have learned to say such sentences when being in this neurological state, and if the conditions obtain which induce me to say the sentences). Moreover, if I *know* that I have a pain, or if I am *aware* of the pain, then again I am merely in a particular neurological state, which may or may not be different from the previous one. Finally, if, according to mentalist views, I am supposed to be using my (mental) free will when saying, 'I know that I am in pain', then, according to the neurological hypothesis, I am again in a particular neurological state, which may or may not induce me to say the sentence. And in order to individuate these states, I make use of all available data, including those used by mentalists for individuating mental states of free will.

This shows that the neurological hypothesis replaces not only the hypothesis of *other* minds, but of minds in *general*, including my own. And our discussion in Section IV clearly shows that the neurological hypothesis, which is obtained by applying the method of Section III, is better than the general mentalist hypothesis.

It is true that the mentalist hypothesis is more intuitive for us. But this may be because to prefer a theory, whose only virtue is that it is simpler than alternative theories, is perhaps not always an intuitive attitude. For many people, it was more intuitive to believe that lightning and thunder were caused by Zeus' thunderbolt than to attribute them to natural causes, or to accept Ptolemy's astronomy rather than Copernicus' theory, even when the latter theory was already available. There may be many reasons which give an hypothesis an intuitive character. But if we apply the methodological criteria accepted by most modern scientists, then we have to conclude that the mentalist hypothesis is not the best explanation of the relevant evidence. It is more complex than the neurological hypothesis, yet its greater complexity is not compensated by an increase in explanatory and predictive power.[10]

NOTES

[1] We shall see that the hypothesis also abstains from assuming the existence of my own mind, and it suggests a different terminology for speaking about my mental experiences.

[2] In Section V, I analyze in more detail the verbal reports of mental states.

[3] Some of these procedures reflect interactionist and epiphenomenalist views.

[4] If our natural language contains expressions which refer to certain mental states but no expressions which refer to the corresponding neurological states, then we can still make use of the mental expressions. We can employ them for naming the neurological states, by giving them a neurological, instead of a mental interpretation. (In Stemmer, 1983, I adopted a somewhat different approach. See especially, pp. 27f and fn. 17.)

[5] There will be such remaining mental states only if it is assumed that certain fine differences between mental states – e.g., between the state corresponding to the belief that all ravens are black and the one corresponding to the belief that all ravens are grey – are not reflected by neurological differences. Although it is unlikely that mentalists would wish to adopt this position, I do not want to exclude this possibility.

[6] If we adopt the procedure mentioned in N. 4, and if we normally use 'pain in the hand' for referring to state *a*, then we can continue to say that the man has pain in the hand. This now means that he is in the neurological state *d*.

[7] Mentalists, with the help of observational data, are frequently able to distinguish not only between individual mental states of a person, such as between his headache yesterday and his headache today, but also between types of mental states of a particular person, such as between his headaches and stomachaches. Similar distinctions can often also be made for the mental states of people in general, and perhaps of other organisms as well. We notice that the neurological hypothesis can make analogous distinctions between neurological states, since it makes also use of these observational data.

[8] There are additional criteria that one might consider for theory choice. But I have preferred to concentrate here on simplicity and observational content, since they seem to be the most important. Still, let me point out that the neurological hypothesis is also favored by the criterion which recommends coherence and continuity with well-established theories. The neurological entities which are assumed by the hypothesis have the materialistic characteristics of the entities that are assumed by most well-established

theories. On the other hand, there are no, or at most very few, well-established theories which assume entities having the special characteristics of mental entities.

[9] See, e.g., Quine's (1974) treatment of language acquisition.

[10] I wish to thank Alex Blum for helpful comments on an earlier version of this paper.

BIBLIOGRAPHY

Churchland, P. M.: 1981, 'Eliminative materialism and propositional attitudes', Journal of Philosophy 78, pp. 67–90.

Conee, E.: 1985, 'Physicalism and phenomenal qualities', The Philosophical Quarterly 35, pp. 297–302.

Jackson, F.: 1982, 'Epiphenomenal qualia', The Philosophical Quarterly 32, pp. 127–136.

Pargetter, R.: 1984, 'The scientific inference to other minds', Australasian Journal of Philosophy 62, pp. 158–163.

Price, H. H.: 1960, 'Some objections to behaviorism', in S. Hook (ed.), Dimensions of Mind (Collier Books, New York), pp. 79–84.

Quine, W. V.: 1974, The Roots of Reference (Open Court, La Salle, Ill.)

Stemmer, N.: 1983, The Roots of Knowledge (St. Martin's Press, New York).

Department of Philosophy,
Bar-Ilan University,
Ramat-Gan,
Israel.