KRIPKE ON THE IDENTITY THEORY *

AUL Kripke has presented a number of arguments against various versions of the psychophysical identity theory.† It is not entirely clear that Kripke wants to endorse every one of these arguments, but many of his readers have been impressed by them, some even to the point of thinking them to be conclusive refutations of the views in question. Thus, the arguments deserve careful attention, whether Kripke accepts them or not.

These arguments are directed against at least three different sorts of psychophysical identity thesis.¹ As I understand them, the three main claims Kripke wants to deal with are: first, the claim that each person is identical to his or her body; second, the claim that each particular mental event or state is identical to some corresponding particular physical event or state; and third, the claim that each "type" of mental state is identical to some corresponding "type" of physical state.

Some of Kripke's readers have concluded that what makes his arguments especially noteworthy is that, in them, Kripke has made essential use of some novel and impressive technical machinery. The concept of the rigid designator, and the principle of the necessity of true identity statements in which rigid designators appear, are the primary technical features in question. Kripke himself says that these "analytical tools go against the identity theory" (I&N 163).

I have two main purposes in the present paper. First, I want to present and evaluate Kripke's arguments against the first and second versions of the identity theory. I do not understand the third version well enough to know whether anyone ever held it, or whether Kripke's arguments have any bearing on it. So I won't discuss them here. I will try to show, with respect to the arguments I do consider, that every one of them is open to serious objection. My second main aim is to determine the role in these arguments of the concept of the rigid designator, as well as that of Kripke's principle. I will try to show that Kripke's principle, whose truth

^{*} To be presented in an APA symposium on Kripke and Materialism, December 29, 1974. William G. Lycan will be co-symposiast; see his "Kripke and the Materialists," this JOURNAL, this issue, 677–689. Diana Ackerman will comment; her comments are not available at this time.

t "Identity and Necessity," in M. Munitz, ed., *Identity and Individuation* (New York: NYU Press, 1971), pp. 135–164 (henceforth I&N); and "Naming and Necessity," in D. Davidson and G. Harman, eds., *Semantics of National Language* (Dordrecht: Reidel, 1972), pp. 253–355 (henceforth, N&N). It should be noted that these papers are transcripts of lectures given without notes, and so one cannot fault Kripke if in some places the arguments seem to lack rigor.

¹ Kripke makes the distinction at N&N 334.

is not here in question, cuts no ice with respect to the identity theory.

Let us consider first what Kripke says against the first sort of claim, the claim that each person is identical to his or her body.

1

Kripke suggests that the person-body identity theory can be refuted "without the use of a modal argument" (I&N 164, fn 19). In "Naming and Necessity" the refutation is stated rather briefly:

Of course, the body *does* exist without the mind and presumably without the person, when the body is a corpse. This consideration, if accepted, would already show that a person and his body are distinct.²

Surely Kripke does not mean to suggest that, in every case, the body continues to exist after the person dies. The counterexamples to that claim may be somewhat gruesome, but are sufficiently obvious. So Kripke's point must be that in some cases the body, but not the person, continues to exist after death. This premise, apparently, is supposed to imply that, in those cases at least, the person and the body are diverse.

So the argument can be formulated as follows, letting Pxt', Ext', Bxt', and Oxyt' mean respectively, "x is a person at t," "x exists at t," "x is a body at t," and "x belongs to y at t":

ARGUMENT A

- (1) $(\exists x)$ $(\exists y)$ (Pxt & Ext & Byt & Eyt & Oyxt & Eyt' & ~Ext')
- (2) (x) (y) (Ext' & $\sim Eyt' \supset x \neq y$)
- (3) $(\exists x)$ $(\exists y)$ $(Pxt & Byt & Oyxt & x \neq y)$

Kripke presents a very similar argument in another place. There he formulates the argument as follows, letting 'A' be a name for Descartes's body:

... provided that Descartes is regarded as having ceased to exist upon his death, "Descartes $\neq A$ " can be established without the use of a modal argument; for if so, no doubt A survived Descartes when A was a corpse. Thus A had a property (existing at a certain time) which Descartes did not (I&N 164, fn 19).

This argument can be rephrased as follows, using 'd' and 'b' as names for Descartes and his body respectively, and 't' as a name for some time shortly after Descartes's death:

ARGUMENT B

- (4) Ebt & ~Edt
- (5) (x) (y) (Ext & $\sim Eyt \supset x \neq y$)
- (6) $b \neq d$
- 2 N&N 354, fn. 73. It is not clear that Kripke endorses this argument.

Arguments A and B are valid. Furthermore, in each case the conclusion is inconsistent with person-body materialism. Nevertheless, neither argument is very persuasive: in each case anyone predisposed to believe in person-body materialism would be right to claim that the argument either begs the question or else is based on a failure to understand what personal-body materialism amounts to.

Person-body materialism is the doctrine that people are their bodies. In other words, anything that is a person is identical to the body of that person:

(7)
$$(x)$$
 (t) $(Pxt \supset (\exists y) (Byt & Oyxt & x = y))$

Any clear-headed proponent of this view would undoubtedly say that death usually does not make people cease to exist. Since, on this view, people are their bodies, the conditions of survival of any person must be the same as those of his or her body. This is not to say that anyone adopting materialism would be saddled with the view that there must be some sort of life after death. Far from it; this form of materialism doesn't even imply that people continue being people after they die. But what's important here is that, on this view, the property of being alive is an accidental property of the things that have it. Something can have it for a while, and then cease to have it, without thereby ceasing to exist. Thus, the things that are people can continue to exist after death.

The problem with Kripke's arguments, then, is that in each case the argument seems to presuppose that at death the person literally ceases to exist, whether the body does or not. In argument A this presupposition appears as (1). In B, it is (4). Anyone wishing to defend person-body materialism should deny these premises. In the case of (1), he should say that, whether it is still alive or not, the thing that was the person at t continues to exist at t. This follows from the fact that, on this view, the thing that was the person is the thing that was the body. According to (1), that thing still exists at t. In the case of (4), he should say that, since b still exists at t, so does d. Of course, such materialists will have to grant that, when he is dead, Descartes is no longer a living person, and may even have ceased to be a person altogether.

Perhaps Kripke or someone wishing to defend these arguments would say that the property of being alive is an essential property of the things that have it. We can understand this to mean that everything that is alive is such that, necessarily, for any time, t, if it exists at t, then it is alive at t:

(8)
$$(x)(t)(Axt \supset \Box(t')(Ext' \supset Axt'))$$

From this it would follow that, when Descartes died, he literally ceased to exist. This might be thought to provide support for (4).

However, the claim (8) that being alive is an essential property of the things that have it does not provide support for (4). Indeed, together with some pretty obvious facts, it entails the denial of (4). Surely, Descartes's body was alive when Descartes was. If being alive is essential to the things that have it, then, just as death marked the end of Descartes, so must it have marked the end of his body as well. On this view the corpse must be construed as a new entity, brought into existence by "substantial change." Hence, far from providing support for it, the claim that everything that is alive is essentially so actually lends overwhelming support to the denial of (4).

The Kripkean thus has to say, if he wants to adopt this line of defense for (4), that being alive is an essential property of Descartes, but not an essential property of his body, even though it is a property his body has.³ Hence, the body can survive death, while Descartes cannot. This view might be thought to lend some credibility to the claim that, after death, Descartes's body continues to exist, but Descartes does not.

There are a number of points to notice about this way of defending (4). First, it makes argument B as a whole superfluous, since the defense of (4), together with Leibniz's law, implies the desired conclusion. The argument would be:

ARGUMENT C 4

- (9) d has the property of being essentially alive.
- (10) b does not have the property of being essentially alive.
- (11) x = y iff (F) (Fx iff Fy)
- (12) $d \neq b$

A second point to notice is that it is no longer clear that the argument, if developed in this way, can be used to show that the person-body identity theory can be refuted "without the use of a modal argument." For the claims about the essential properties will be understood, by some at least, to be modal claims.

Third, we should note that no argument has been given for (9). The person-body materialist, thinking as he does that people are their bodies, will undoubtedly reject this premise. For, on his view, no person is essentially alive. He finds abundant evidence for this

³ It might be thought that there can be no such essential properties. But consider the property of being either Nixon or snub-nosed. Some things have this property accidentally, and other have it essentially.

⁴ This argument is valid whether 'd' and 'b' are rigid or not.

in the fact that there are so many corpses, every one of which proves (at least to materialists) that something that is a person can be alive at one moment, and fail to be alive at the next. Hence, nothing that is a person is essentially alive. If a Kripkean assumes that each person is essentially alive, then he assumes something none of his philosophical opponents should assume, and something for which Kripke has given us no good argument. So it appears that neither argument A nor argument B should persuade anyone inclined toward person-body materialism. Each employs an undefended premise no materialist should accept.

Finally, we should note that neither argument makes use of the concept of the rigid designator or of Kripke's principle of the necessity of true identity statements in which rigid designations appear.

Let us now turn to a somewhat more interesting argument Kripke has presented against the first form of materialism. Kripke ascribes this argument to "Descartes and others following him" (N&N 334). It is based on the idea that each person can exist without his or her body. As Kripke puts it:

Let 'Descartes' be a name, or rigid designator, of a certain person, and let 'B' be a rigid designator of his body. Then if Descartes were indeed identical to B, the supposed identity, being an identity between two rigid designators, would be necessary, and Descartes could not exist without B and B could not exist without Descartes (N&N 334).

I think the argument can be rephrased as follows:

ARGUMENT D

- (13) 'd' is rigid and 'b' is rigid.
- (14) ' \diamondsuit (3t) (Edt & \sim Ebt)' is true.
- (15) If α is rigid, and β is rigid, and $\lceil \lozenge \ (\exists t) \ (E\alpha t \& \sim E\beta t) \rceil$ is true, then $\lceil \lozenge \ (\alpha \neq \beta) \rceil$ is true.
- (16) If α is rigid, and β is rigid, and $\lceil \lozenge \mid (\alpha \neq \beta) \rceil$ is true, then $\lceil \alpha \neq \beta \rceil$ is true
- (17) ' $d \neq b$ ' is true.

Kripke remarks, in discussion of this argument, that "a philosopher who wishes to refute the Cartesian conclusion must refute

⁵ The person-body materialist, noting that some corpses are now not alive, but formerly were people, will deduce:

$$(\exists x)$$
 $(\exists t)$ $(\exists t')$ $(Pxt & Ext' & \sim Axt')$

This is inconsistent with the claim that every person is essentially alive:

$$(x)$$
 (t) $(Pxt \supset \Box (t') (Ext' \supset Axt'))$

the Cartesian premise, and the latter task is not trivial" (N&N 335). Kripke is right about this, but it should be equally obvious that a philosopher who wishes to make use of this argument to establish the Cartesian conclusion must first establish the Cartesian premise. That is, he must first establish (14). And that task is equally nontrivial.

It might be thought that, since his target is the so-called "contingent identity theory," Kripke is justified in assuming (14) for the purposes of his argument. But this would be a mistake. The contingent person-body identity theory does not entail anything like (14).

The contingent person-body identity theory is the view that each person is identical to his or her body, and, furthermore, that it is not necessary that this be so. It is true only as a result of physically necessary, but not metaphysically necessary laws of nature. In other words, the view Kripke wants to refute can be formulated as:

(18) (x) (t)
$$(Pxt \supset (\exists y) (Byt & Oyxt & x = y))$$

 $& \sim \square(x) (t) (Pxt \supset (\exists y) (Byt & Oyxt & x = y))$

(18) does not imply that each person and his or her body are such that it is only a contingent fact that they are identical. To get that doctrine, we would have to shift the position of the modal operator in (18), so as to make it affirm "de re" contingency, as in:

$$(19) (x) (x) (t) (Pxt \supset (\exists y) (Byt & Oyxt & x = y & \sim \square (x = y))$$

(19), perhaps, would give us the conclusion that it is possible for a person and his body to fail to coexist. In other words, given (19), statements like (14) might be more plausible.

But it should be obvious that (18) does not imply (19). For (18) would be true if every person in this world were identical to his or her body, although in other possible worlds there were people who are "pure spirits." That would not suffice to make (19) true. Indeed, so long as there are people, nothing could make (19) true.

Thus, nothing in the contingent-identity theory has as a consequence the view, expressed by (14), that Descartes and his body could have failed to coexist. So the burden is on Kripke to give us some reason to accept this controversial premise. What sort of defense might be given? It could be argued that incorporeal survival of death is at least conceivable, and so it is possible for Descartes to exist without any bodies. In other words, it seems to be possible for there to be a time at which Descartes exists, but at which no bodies exist. Furthermore, Descartes's body is a body. Therefore,

it could be argued, it is possible for there to be a time at which Descartes exists, but at which his body does not exist. Hence, we have a proof of (14).

This argument is sophistical, and its invalidity should be made clear. Using the abbreviations already introduced, we can reformulate it as:

ARGUMENT E

- (20) \Diamond ($\exists t$) ($Edt \& \sim (\exists x) (Ext \& Bxt)$)
- (21) (t) $(Ebt \supset Bbt)$
- (22) \Diamond ($\exists t$) ($Edt \& \sim Ebt$)

Even though 'd' and 'b' are rigid, the argument as a whole is invalid. This can be seen by comparing it with a formally parallel argument in which the premises are clearly true, and the conclusion clearly false. To do so, reinterpret argument E as follows: let 'd' refer to Cicero, 'b' to Tully, and let 'Bxt' mean "x is a Roman at t." Surely it does not follow from the fact that Cicero could have existed at a time when there were no Romans, and the fact that Tully was a Roman, that there could have been a time at which Cicero existed without Tully.⁶

One way to make the argument valid would be to strengthen the second premise. Instead of (21), we could have:

(23)
$$\square$$
 (t) (Ebt \supset Bbt)

In this case, the second premise says, in effect, that b is essentially a body—it is necessary that b be a body whenever it exists.

The revised argument, which we can call "argument F," is valid, but it certainly should not convince any materialists. Some materialists may believe that every body is essentially a body. They will claim that, since Descartes is a body, and every body is essentially a body, it follows that Descartes is essentially a body. From this they will infer that it is impossible for there to be a time at which Descartes exists, but at which no bodies do. Thus, they will say that (20) is false.

Other materialists, who reject the idea that every body is essentially a body, will point to (23) as the weak link. Thus, no materialist should be convinced by argument F.

A more interesting point about argument F is that, if we use it to buttress (14) of argument D, we make argument D as a whole pointless. For (20) and (23) independently entail that Descartes is diverse

⁶ I assume that the property of being Roman is an accidental property of Cicero. If the example is unacceptable, select some property Cicero had throughout his existence, but which he could have lacked.

from his body. (20) says, in effect, that Descartes could have existed even if there were no bodies. (23) says, in effect, that Descartes's body could not have existed if there were no bodies. Thus, there is a property, possibly existing at a time when no bodies exist, that Descartes is alleged to have and that his body is alleged to lack. By Leibniz's law, they must be distinct. Thus, as in arguments A and B, it seems that an undefended and controversial essentialism is what entails dualism. Given this essentialism, argument D, in which Kripke's principle appears, is superfluous. Without the essentialism, argument D contains a premise no materialist has any reason to accept.

So it appears that Kripke hasn't produced any persuasive argument against the first version of the identity theory. Furthermore, it appears that his principle of the necessity of true identities with rigid designators plays no important role in the arguments he has produced. Let us move on, then, to a consideration of Kripke's arguments against the second version of the identity theory.

TT

In "Naming and Necessity" Kripke presents two closely related arguments against the view that each particular mental event is to be identified with a particular brain event. These arguments have a good deal in common, and the bulk of my comments are not directed toward anything in this common part. Thus, it may be convenient to isolate this part, so as to avoid repetition. We can represent it as follows:

ARGUMENT G

- (24) 'a' and 'b' are rigid designators for some pain-event and the corresponding brain-event.
- (25) ' \Diamond ($a \neq b$)' is true.
- (26) If α is rigid, and β is rigid, and $\lceil \lozenge (\alpha \neq \beta) \rceil$ is true, then $\lceil \alpha \neq \beta \rceil$ is true.
- (27) ' $a \neq b$ ' is true.

The argument proceeds from the possible diversity of a and b, via Kripke's principle of the necessity of true identities with rigid designators, to the conclusion that a, the pain-event, is diverse from b, the brain-event. As should be obvious, the crucial premise is (25), the claim that the pain-event and the brain-event are possibly diverse.

It should be noted again, perhaps, that, in any coherent form, the contingent pain/brain-event identity theory does not entail (25).

⁷ I shall not discuss the corresponding passages in I&N.

From the fact that it is contingently true that every pain-event is identical to some brain-event, it does not follow that every pain-event is such that there is some brain-event such that it is contingent that the former is identical to the latter. Thus, Kripke needs to defend (25), and he attempts, in at least two different ways, to do so.

In one argument, (25) is defended by appeal to the alleged fact that it is logically possible for b, the brain-event, to occur without "Jones feeling any pain at all, and thus without the presence of [a]" (N&N 335). In the other argument, (25) is defended by appeal to the alleged fact that "the pain could have existed without the corresponding brain state" (N&N 336). Let us consider the first argument more closely.

Kripke writes:

Prima facie, it would seem that it is at least logically possible that b should have existed (Jones's brain could have been in exactly that state at the time in question) without Jones feeling any pain at all, and thus without the presence of [a]. Once again, the identity theorist cannot admit the possibility cheerfully and proceed from there; consistency, and the principle of the necessity of [true] identit [i] es using rigid designators disallows any such course (N&N 335).

It may appear that Kripke has his opponents in a difficult spot. After all, it is allegedly only in virtue of contingent laws of nature that we feel pain when such brain-events occur. Thus, Kripke seems to be right in saying that it is possible for b, the brain-event, to occur without anyone being in pain. From this, and the fact that a is a pain-event, it seems to follow that b could have occurred without a. But surely if b could have occurred without a, b could have been diverse from a. Thus, we seem to have our proof of (25).

The argument can be reformulated as follows, using 'Oxt' and 'Pxt' to mean "x occurs at t" and "x is a pain-event at t."

```
ARGUMENT H
(28) \Diamond (\exists t) (Obt \& \sim (Ex) (Pxt \& Oxt))
(29) (t) (Oat \supset Pat)
(30) \Diamond (\exists t) (Obt \& \sim Oat)
(31) \Diamond (\exists t) (Oxt \& \sim Oyt) \supset \Diamond (x \neq y)
(24) \Diamond (a \neq b)
(30), (31)
```

Argument H is not valid. For (28) and (29) do not entail (30). The problem here is formally analogous to the problem discussed earlier in connection with argument E (671 above). It does not follow, from the fact that there is a possible world in which b

occurs without anything that is there a pain-event, and the fact that a is a pain-event in this world, that there is a world in which b occurs without a. For if a, which is a pain-event in this world, occurs in other worlds without being a pain-event, then perhaps a does occur wherever b does. So the argument turns on the question whether a, which is a pain-event in the actual world, is a pain-event in every world in which it occurs. In other words, the argument turns on the question whether a is essentially a pain-event.

We can bring out this fact by noting that if we change (29) to

(29')
$$\Box$$
 (t) (Oat \supset Pat)

then the revised argument, which we can call argument H', is valid. None of this, of course, is news to Kripke. His defense of (29') is spirited:

The difficulty can hardly be evaded by arguing that ... being a pain is merely a contingent property of a, and that therefore the presence of b without pain does not imply the presence of b without a. Can any case of essence be more obvious than the fact that being a pain is a necessary property of each pain? ... Consider a particular pain, or other sensation, that you once had. Do you find it at all plausible that that very sensation could have existed without being a sensation, the way a certain inventor (Franklin) could have existed without being an inventor (N&N 335).

So Kripke's strategy is clear. He means to make use of something like argument H' to support (25). (25), together with the rest of argument G, entails the denial of the pain/brain-event identity theory.

There are a number of points to notice about this way of arguing for (25). First of all, we should note that the premises of the argument for (25) independently entail the denial of pain/brain-event identity theory. For (28) says, in effect, that the brain-event b has the property of possibly occurring when no pain-event does. (29') says, in effect, that a, the pain-event, has the property of necessarily occurring only when a pain-event occurs. In other words, a lacks the property of possibly occurring when no pain-event occurs. A straight-forward application of Leibniz's law entails that a is diverse from b. Argument a, in which Kripke's principle appears, is superfluous.

So Kripke's principle of the necessity of true identities with rigid designators plays no important role in the argument. If the essentialistic premises, (28) and (29'), are correct, we can validly infer the conclusion without making use of the principle. On the

other hand, if the essentialistic premises are not correct, then we have lost the defense of (25), and so there is nothing to which we can apply Kripke's principle.

A second interesting feature of this defense of (25) is that, aside from using rhetorical questions to elicit our agreement, Kripke nowhere attempts to establish the claim that every pain is essentially a pain. He asks whether any case of essence could be "more obvious than the fact that being a pain is a necessary property of each pain?" (N&N 335) Elsewhere he urges us to consider some pain we have felt and ask ourselves whether "it is possible that it itself could have existed, yet not have been a pain?" (162) He suggests that, in his opinion, it would be "self-evidently absurd" (336) to say that something that is in fact a pain-event could have occurred without being a pain-event. Obviously, none of this constitutes a reasoned defense of the doctrine in question.

Kripke, and perhaps others who have thought about it, may have felt that (29') is a premise so basic that all reasonable parties to the dispute will agree to it. Kripke's strategy may be to show that, having admitted (29') we are committed to a rejection of the pain/brain-event identity theory.

To view the matter in this way strikes me as naive. For any serious materialist should recognize that his view entails that painfulness is never part of the essence of a pain-event. Pain-events are experienced as they are only as a result of contingent laws of nature. Early materialists may have had just this point in mind when they asserted, somewhat confusedly in my opinion, that certain brainevents are "contingently identical to" pain-events. Understood charitably, what they must have meant is that it is contingently true that brain-events of certain descriptions are identical to pain-events. These very brain-events, had the laws of nature been different, would of course still have been self-identical, but would not have been identical to anything that would, under those circumstances, have been a pain-event. Thus, such events are not essentially painful. The very same point can be put more straightforwardly by saying that certain brain-events are such that it is contingent that they are felt as pains. This way of putting the view is less confusing, since we don't speak of "contingent identity," but the point remains the same.

In light of this, it appears to me that (29') has the following features. First, without it, argument H is not valid. With it, argument H' is valid, but then argument G is superfluous. For, together with (28) and Leibniz's law, (29') entails the denial of the pain/

brain-event identity theory. Second, although Kripke steadfastly affirms the truth of (29'), he nowhere argues for it. Finally, Kripke's philosophical opponents have no reason to accept (29'), and, in some cases at least, have meant to reject it from the outset. Thus, it appears that argument H' turns on an undefended, controversial premise that materialists have, and should have, rejected. (25) remains unsubstantiated, and we have so far been given no good reason to be moved by argument G.

Kripke has at least one other way of defending (25). Instead of arguing from the premise that it is possible for b to occur without anyone being in pain, he seems to argue from the premise that, whereas it is impossible for b to occur without certain brain cells being in a certain configuration, it is possible for a to thus occur (N&N 336). Space limitations prevent a full discussion of this argument, but it appears that its defects are similar to those of argument H'. The essentialistic premises entail the diversity of a and b without appeal to Kripke's principle; Kripke nowhere defends the essentialistic premises; and they would surely (and plausibly) be rejected by any clear-headed materialist.

So neither of Kripke's arguments against the second form of the identity theory is persuasive. Contingent-identity theorists of the relevant kind, if there are any, have been given no reason to fear that their view, when properly formulated, is undermined by Kripke's analytical tools.

m

What conclusions can be drawn from this survey? First, I think it should be clear that there is no truth in the idea that Kripke's analytical tools "go against the identity theory" in either of the forms we have considered. The concept of the rigid designator is undoubtedly of considerable value in the philosophy of language or in modal logic. But it has no direct bearing on the truth or falsehood of the identity theory.

Second, none of Kripke's arguments breaks new ground in the debate about the identity theory. In every case, the argument depends upon some premise that defenders of the identity theory have been given no reason to accept, and in many cases these are premises they have already rejected.

Finally, I think we should agree with Kripke's suggestion that, in spite of his arguments, the issue is still "wide open and extremely confusing."

FRED FELDMAN