MARK E. KALDERON

EPIPHENOMENALISM AND CONTENT

(Received 7 August, 1986)

A salient feature of any given explanation is that it should have a certain modal force — that it should be able to support the appropriate sorts of counterfactuals and subjunctive conditionals. And so it is with action explanation. Rationalizing explanations make essential use of the content of the agent's propositional attitudes. Thus if an agent desires that p and believes that, by doing A, p will be realized, the agent will perform A. If this familiar form of reasoning is to have any explanatory bite, then it must be able to support certain counterfactuals. We want to be able to say that if the agent didn't have the relevant beliefs and desires, he wouldn't have performed action A.

Token physicalism purports to resolve the problem about psychophysical causation by reducing it to the relatively unproblematic physical-physical causation. Thus a physical event p_1, which is token identical with a contentful mental event, causes a physical event p_2, which is itself token identical with an intentional action. According to Davidson, a singular causal statement will be subsumed by some physical law relative to the appropriate descriptions. But the causal or nomological necessity attached to the singular causal statement in virtue of its nomic subsumption isn't sufficient to capture the appropriate counterfactuals required of our belief-desire explanations. I want to argue that if, as Davidson claims, rationalizing explanations are a species of causal explanation, then it isn't enough that the relevant propositional attitudes are token identical with the appropriate physical events, or even that the content of the relevant propositional attitudes "weakly supervene"[1] on the physical state of the agent, but rather, in order to capture the modal force of folk psychological explanations, the relevant propositional attitudes must have their causal efficacy in virtue of their semantic content. That is to say that having semantic content, being a belief that p, must be a causally relevant property. If, on

Philosophical Studies **52** (1987) 71—90.

Davidson's account, the content of the relevant propositional attitudes fail to be causally relevant to the production of the given intentional action, then, *ipso facto*, his notion of the mental is epiphenomenal.[2] I suspect that these considerations apply *mutatis mutandis*, to any given autonomist theory — that the threat of epiphenomenalism is a general problem for any non-reductionist theory of content.[3]

COUNTERFACTUALS AND EXPLANATION

Davidson's project in "Actions, Reason's and Causes"[4] is to provide an analysis of the relation between a reason and an action in virtue of which the reason *explains* the action. He wants to provide an account of what makes rationalizing explanations adequate — how is it that they explain anything at all? Davidson begins by drawing a distinction between an agent merely having a good reason to perform a given action and an agent having a "reason for which" the action was performed. A reason justifies the agent's action in terms of the semantic content of his propositional attitudes, i.e., his beliefs, fears, wishes, hopes, etc. An agent can have many reasons which justify a given action. For instance, one reason for going to bed would be that the agent believes that he needs ten hours of sleep a night and that he has to be up at a certain time. The agent may also be exhausted and thus desire sleep. However, he went to bed because he was exhausted and not because he ought to, the agent being in general irresponsible as to his morning appointments. Davidson claims that a given reason *rationalizes*, or explains, a given action when it is the *reason for which* the action was performed. Rationalizing explanations have their explanatory power by adverting to such reasons, i.e., rationalizing explanations are adequate just in case the reason cited is the reason for which the action was performed.

To see this, simply consider offering as an explanation for the agent's going to bed that he believed that he needed the appropriate amount of sleep. Such an explanation would be inadequate since if he hadn't been exhausted, he wouldn't have gone to bed — regardless of his belief that he ought to. It seems clear that such a reason doesn't adequately explain the agent's action since, in effect, it won't support the appropriate sorts of counterfactuals and subjunctive conditionals. A neces-

sary condition for a given explanation to be adequate, then, is that it should support counterfactuals of the form:

(1) If the agent hadn't had reason R, then he wouldn't have performed action A.

Davidson wants to account for this distinction by proposing that a reason for which a given action is performed is the *cause* of the given action. Let's take a step back and unpack this suggestion. One principal difference between a reason and a reason for which is that the reason for which is a "determinant" of that action. A necessary condition for one event to be a determinant of another is that it should support counterfactuals of the form:

(2) If event e_1 determines event e_2, then if e_1 hadn't obtained, e_2 wouldn't have obtained.

Without such a necessary condition one simply doesn't have an interesting notion of determination. Davidson's proposal is that in order to get the counterfactuals necessary for explanation, a reason must be a causal determinant of the given intentional action. The general strategy is to ground the counterfactuals needed for explanation in the modal properties characteristic of the class of determiner relations.

As a matter of historical fact, given Davidson's extensional framework and his agnosticism towards counterfactuals, it probably isn't true that Davidson was worried about counterfactuals in drawing the distinction between a mere reason and a reason for which. In fact he only mentions counterfactuals as a standard move in distinguishing laws from contingent empirical generalizations. It is clear, however, that he saw the principal difference between the two as grounded in the fact that one was, and the other wasn't, a causal determinant of the given intentional action. The question which is crucially glossed in Davidson's account is why being a causal determinant is an interesting explanatory property. Thus the need to ground the appropriate counterfactuals remains tacit. I plead innocent to the charge of interpretive violence on the basis of the principle of charity — I simply don't know how to make sense of Davidson's answer without unpacking counterfactuals and their role in explanation.

Whether or not my interpretation of Davidson's move is justified,

there are independent reasons for thinking that Davidson's account is only adequate if it can capture counterfactuals like (1) over and above the observation that explanations in general should have a certain modal force — that supporting counterfactuals like (1) is in some sense constituitive of being explanatory. Recall that Davidson claims that rationalizing explanations are a species of *causal* explanation. It follows, then, that the propositional attitudes adverted to in a rationalizing explanation must be among the causal determinants of the given intentional action.

Consider the following argument suggested by Dretske.[5] If propositional attitudes are to be among the causal determinants of a given action, being a belief that *p*, having *that* semantic content as opposed to another, should be a causally relevant property. At least a necessary condition for being a causally relevant property can be spelled out in terms of being able to support the appropriate sorts of counterfactuals and subjunctive conditionals. In particular, causal relevance has the following counterfactual property:

(3) If property P is causally relevant to event e_1 causing event e_2, then if e_1 hadn't exemplified P, e_1 wouldn't have caused e_2.

If the appropriate propositional attitudes having the content that they do is to be causally relevant to the production of the given action, then the semantic content of the pertinent attitudes should support the appropriate counterfactuals. If they fail to do so, they violate this necessary condition and thus aren't causally relevant. Since rationalizing explanations typically advert to the content of propositional attitudes as among the causal determinants of a given action, the semantic content of these attitudes must support the appropriate counterfactuals if one's intentional psychology is to be an adequate explanatory edifice. If rationalizing explanations fail to do so on Davidson's account, then Davidson's notion of content is epiphenomenal.

The above argument is valid. One might object, however, to its soundness without further qualifications. Someone might claim that Dretske has overlooked The Problem of Spurious Overdetermination, e.g., that there might be two mental events which are both independently causally sufficient for the production of a given action. Let m_1 and m_2 be two such events which cause the intentional action A. Since

m_1 and m_2 are independently causally sufficient for A, counterfactuals of the form:

(4) If m_1 hadn't obtained, A wouldn't have been performed

might fail since at some nearby world where m_1 fails to obtain, m_2 obtains. Since m_2 is causally sufficient for A, A will obtain while m_1 doesn't, thus falsifying the subjunctive conditional. It's important to note that this is a general problem — that The Problem of Spurious Overdetermination is not specific to mental causation, and belies, therefore, neither the soundness of Dretske's argument nor the modal intuition behind it.

It's important to note that Dretske's counterfactual condition is just a special case of the proposed necessary condition on the class of determiner relations. Recall that Davidson claims that only reasons which causally determine the given intentional action will be explanatorily adequate. I argued that what makes a mental event's being a determiner of a given action an explanatorily interesting property is that one can ground the counterfactuals needed for explanation in the necessary condition on determiner relations, i.e., that any member of the class of determiner relations should support counterfactuals of the form if event e_1 determines event e_2, then if e_1 hadn't obtained, e_2 wouldn't have obtained. Given Davidson's base distinction, and given that my inference to the best explanation is correct, i.e., that you can't make sense of Davidson's proposal without invoking counterfactuals, then Davidson is committed to (3). Thus if Davidson's theory fails to meet Dretske's condition, it is, on its own account explanatorily inadequate.[6]

DAVIDSON' CAUSAL THEORY OF ACTION

Davidson proposes that the explanatory relation that holds between a reason R an an intentional action A is fully specified by the justificatory relation and the appropriate causal relation. He holds, moreover, that one only understands a given action explanation if one understands how to construct for it a primary reason. A primary reason is a coordinated pair of mental states consisting of a belief-state and a pro-attitude. A pro-attitude is a propositional attitude which expresses a

favorable attitude of the agent towards actions of a certain kind. Such pro-attitudes would include wants, desires, hopes, urges, etc. This then is Davidson's theory:

(5) R is a primary reason why an agent performed the action A, under description D just in case
  (i) R consists of a pro-attitude of the agent towards actions with a certain property and a belief that A, under description D, has that property, and
  (ii) R bears the appropriate causal relation to A.

The justificatory relation arises out of the semantic content of the relevant propositional attitudes. Justification is normative or evaluative in the sense that the intentional action must be reasonable in light of the agent's beliefs and desires. A given action is reasonable in light of the relevant belief desire pair in virtue of a piece of practical reasoning.[7] More specifically, it is reasonable in virtue of the quasi-logical relationship that arises out of the semantic content to the appropriate propositional attitudes. Davidson thinks that an intentional action expresses something like propositional content. This propositional content is an unconditional evaluative proposition that the agent's action is desirable. Davidson holds that the unconditional evaluative proposition is determined by a non-deductive inference whose premisses are the appropriate belief and pro-attitude. The content of a pro-attitude is a *prima facie* evaluative proposition that any act is desirable insofar as it is of a certain kind. The content of the pro-attitude thus concerns act-types. Since the unconditional evaluative proposition expressed by the intentional action is about an act-token, the belief premiss must itself be about an act-token. Specifically, the agent believes that this token of this act-type is also of the type specified by the pro-attitude.[8] Consider an agent who desires to soothe his nerves. He believes that this act of drinking a shot of whiskey will soothe his nerves, therefore this particular act of drinking a shot of whiskey is desirable. What's important for the purposes of this paper is simply that the justificatory relation is systematically related to the content of the agent's primary reason and the unconditional evaluative proposition expressed by the intentional action.

It's important to note that the semantic content of the relevant

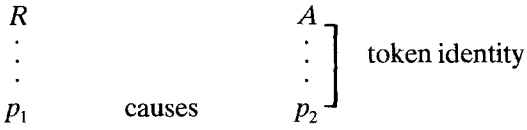
propositional attitudes, while necessary and sufficient for establishing the justificatory relation between the agent's coordinated belief-desire complex and the given intentional action, only provides a necessary condition for primary reason R to be a reason for which. In practical reasoning, the content of the belief and pro-attitude can have as their conclusion an unconditional evaluative proposition without the intentional action, which would express it, being performed by the agent. This is more or less the starting point of "Actions, Reasons, and Causes" and Davidson's paper on intending.[9] The semantic content of the agent's mental state, while making a systematic contribution to the propositional content expressed by the act-token, is simply insufficient for actually producting the given intentional action. R must, in addition, bear the appropriate causal relation to A. The following is Davidson's analysis of a reason R being the cause of an action A:

(6) A primary reason R is a cause of an action A just in case
  (i) R is token identical with physical event p_1 and A is token identical with physical event p_2, and
  (ii) p_1 causes p_2 where p_1 and p_2 are nomically subsumed by a strict physical law relative to descriptions $D1$ and $D2$.

Let's begin by looking at Davidson's views concerning laws and causal relations. In "Mental Events",[10] Davidson holds The Principle of the Nomological Character of Causality which states that events related as cause and effect are subsumed by strict laws. For Davidson, a law is a true lawlike sentence, where a sentence is lawlike if and only if it is a generalization confirmable by its positive instances and can support the appropriate counterfactuals and subjunctive conditionals. Sentences are what primarily instantiate laws. A given event is nomologically subsumed only by satisfying a certain description which itself instantiates a component of the law such as an antecedent or consequent. It is clear that Davidson holds the neo-Humean view that every true singular causal statement entails a law, but he argues that this doctrine is systematically ambiguous. On one reading, the singular causal statement entails a particular law. In action theory this would correspond to the Hempelian deductive-nomological model with its explicit use of covering laws connecting attitudes with actions.[11] On the second reading, a

singular causal statement only entails the existence of some subsumptive law. Thus to understand a causal explanation given in terms of a singular causal statement it isn't necessary to have any kind of epistemological access to the underlying strict physical law.

This oblique view of causal explanation in conjunction with a physicalist notion of causation merges nicely with Davidson's token physicalism. Davidson embraces token physicalism as an attempt to resolve the problem about psycho-physical causation. Thus a contentful mental event is causally efficacious by being token identical with a physical event. The idea is that a single event can instantiate both mental and physical types. Thus rationalizing explanations, which make essential use of propositional attitudes, entail the existence of some strict law. Davidson's picture seems to be this:



where reason R is token identical with physical event p_1 ($R =_t p_1$) and intentional action A is token identical with p_2 ($A =_t p_2$) and p_1 is causally sufficient for p_2. "p_1 causes p_2" is subsumed by some strict physical law.

Let's briefly sum up. According to Davidson's theory, the explanatory relation "supervenes" on the justificatory relation and the appropriate causal relation, where the justificatory relation arises in virtue of the semantic content of the relevant propositional attitudes. (6) gives necessary and sufficient conditions for a primary reason to be a cause of a given intentional action — but the explanatory relation is supposed to be determined by the justificatory relation and the *appropriate* causal relation. Davidson never provides an analysis of this restriction, but what he has in mind is problems with wayward or deviant causal chains.

THE CAUSAL IRRELEVANCE OF INTENTIONAL ASCRIPTION

Is Davidson's account adequate to capture his base distinction between a mere reason and a reason for which? Does Davidson's machinery have enough apparatus to ground the counterfactuals he needs?

It's given that we want rationalizing explanations such as the belief-desire explanations typical of our vernacular psychology to support counterfactuals like (1). At this point I want to ask a metaphysical question: in virtue of what do rationalizing explanations have their modal force? Since Davidson gives necessary and sufficient conditions for the explanatory relation, if successful, they should entail counterfactuals like (1). In order for Davidson to provide a non-question begging answer to the metaphysical question he must appeal only to what appears on the right-hand side of the biconditional. This is an important restriction. For instance one couldn't take the following line. Causal explanations have the canonical form of a singular causal statement:

(7) *c* caused *e*

and in principle warrant the following counterfactual:

(8) If *c* had not obtained, *e* wouldn't have obtained.

Since rationalizing explanations are a species of causal explanation they should entail counterfactuals like (1). It should be obvious that this response is blatantly question begging. It's based on the left-hand side of the biconditional given in (6). Since Davidson provides an *analysis* of this he should employ his analysis in answering the metaphysical question. I want to argue that Davidson doesn't in fact provide a non-question begging answer to the metaphysical question, that his account fails to support counterfactuals like (1). The principal worry is similar to one voiced by Stoutland[12] — that there is an insufficient connection between the justificatory and causal relations. The present argument is more focused than Stoutland's in that it shows how in particular the connection is insufficient, i.e., because it fails to support the appropriate sorts of counterfactuals and subjunctive conditionals.[13]

Let's begin with the justificatory relation that is specified in the first conjunct of the biconditional given in (5). Justification is normative in that the given act-token must be reasonable in light of the agent's beliefs and desires. But the justificatory relation won't support counterfactuals like (1). Remember, it is essential to Davidson's base distinction in "Actions, Reasons, and Causes", between a mere reason and reasons for which, that the semantic content of the relevant propositional

attitudes alone, while providing a necessary condition for the production of a given intentional action, isn't a sufficient one. It isn't enough that we have as a conclusion of some practical inference an unconditional evaluative proposition concerning the reasonableness of the given act-token. On Davidson's account, the justificatory relation, by itself, is incapable of supporting counterfactuals like (1).

A primary reason must, in addition, be the cause of the intentional action. The constituent belief state and pro-attitude must be token identical with a physical event p_1 ($R =_t p_1$) which causes physical event p_2 which is, itself, token identical with the appropriate intentional action ($A =_t p_2$). On Davidson's analysis, we have the following singular causal statement nomically subsumed by some strict physical law:

(9) p_1 caused p_2

which supports the following counterfactual:

(10) If p_1 hadn't obtained, then p_2 wouldn't have obtained.

Davidson isn't home yet. From the counterfactual given in (10) he must derive (1). But in order to do so there must be some relation R^* between the propositional attitudes mentioned in (1) and the physical events occuring in (9) and (10). If there is no relation whatsoever between the contentful mental events and the appropriate physical events, then there's no way to get from (10) to (1).

On Davidson's account however, there is some relation, namely token identity. He claims that an event-token of a given intentional type is identical with an event-token of a given physical type. Is token identity sufficient to insure the above inference? It's important to note that token identity is modally flaccid since the event-token only contingently exemplifies the given mental type. It's, in effect, a relation that occurs within a world. In a given causally possible world w $R =_t p_1$. But the token identity that holds between R and p_1 won't hold up when you move across causally possible worlds. Token identity is a contingent relation. It is contingent in virtue of the fact that the mental or intentional type used to pick out the event-token is only contingently instantiated in that event-token. Why is this important? Well it seems that, in order to get from (10) to (1), the relation R^* between R and p_1 and A and p_2 respectively must be a modal relation. R^* must itself be

able to support the appropriate sorts of counterfactuals and subjunctive conditionals.

To see this, let's assume the opposite. Let's assume that R^* can be a contingent relation like token identity and can support counterfactuals like (1). Consider the following causally possible worlds:

w_1 · (p_1 causes p_2) and ($R^*(p_1, R)$ and $R^*(p_2, A)$)
w_2 · (p_1 causes p_2) and ($-R^*(p_1, R)$ and $R^*(p_2, A)$)
w_3 · p_1 and p_2 don't obtain.

w_1 is stipulated by induction hypothesis, w_1 being the actual world in the model. Since R^* is a contingent relation, p_1 and R won't be so related in all the nearby worlds. Thus there is a nearby causally possible world w_2 where p_1 won't bear R^* to R while p_2 bears R^* to A. If p_1 and p_2 are contingent physical events, there must be some world w_3 where p_2 fails to obtain because p_1 one does given the counterfactual in (10). Remember we assumed that the contingent R^* would support counterfactuals and subjunctive conditionals like (1). But such a counterfactual will fail in the given model because of nearby worlds like w_2, thus deriving a contradiction. Thus R^* can't be a contingent relation.

One might object as follows. Look, you've only considered the physical counterfactuals and token identity. The justificatory relation, though incapable by itself of supporting counterfactuals and subjunctive conditionals, will nevertheless impose an important restriction once the counterfactual situation has been stipulated. Specifically, since A occured there must be an appropriate belief-desire pair such that they have as their conclusion in a piece of practical reasoning the propositional content expressed by A — or else A simply isn't an intentional action. If it turns out that, given the justificatory relation, the propositional content expressed by A will determine a particular primary reason R, then worlds like w_2 aren't a clear cut counterexample since there is no R whose content will have as a conclusion the unconditional evaluative proposition expressed by A.

This objection is far too quick. Let's assume that, given the justificatory relation and the unconditional evaluative proposition expressed by A, we can indeed specify the primary reason R. It's simply false that A fails to be an intentional action at w_2. In this case we have at w_2 a case of causal overdetermination — $R =_t p_3$ at w_2 where p_3 as well as p_1

cause p_2. Even so, it's just not the case that the justificatory relation plus the unconditional evaluative proposition will determine the content of a specific primary reason. From this act of drinking a shot of whiskey being reasonable or good you can derive neither the content of the pro-attitude that any act is desirable in so far as it will soothe the agent's nerves — nor the content of the belief state — that this act of drinking a shot of whiskey will soothe the agent's nerves. The objection just seems to be a non-starter.

One might further object that I'm asking the wrong question. Look, the interpretive schema in accordance with the principle of charity gives us causally efficacious rationalizing explanations. It's not clear that we take the physical event p_1 such that $R =_I p_1$ in the actual world to the appropriate counterfactual situation. We want to ask of R — if R had not occured would A have? Remember, however, our original question was metaphysical — in virtue of what does the modal force of rationalizing explanations hold? The principle of charity and the subsequent interpretive schema may give you epistemological access to the appropriate rationalizing explanations, but it's not in virtue of the interpretive schema that these rationalizing explanations support counterfactuals. Davidson has provided an analysis of rationalizing explanations that purports to give necessary and sufficient conditions for the explanatory relation. If it's not in virtue of this analysis, I simply don't know how to make sense of Davidson's talk of oblique causation and token identity. It just strikes me that this response is either question begging, since it fails to restrict itself to the right-hand side of the biconditional, or a bad reading of Davidson, confusing as it does metaphysical and epistemological issues.

Why can't someone take the following line. Look, you presuppose a certain notion of causal relevance and it's in virtue of this notion — that causally relevant properties should support the appropriate sorts of counterfactuals — that your argument runs. Why should Davidson buy into this notion of causal relevance? Why can't it be the case that it simply doesn't matter that Davidson can't secure counterfactuals like (1) since *that's just how mental properties are causally efficacious*? A mental event has the patterns of causal interaction that it does in virtue of the fact that the given event exemplifies a physical type correspond-

ing to some nomologically subsumptive description. Perhaps that's just what makes mental causation distinctively mental.

Besides the obvious concessionary tone of such a reply, there's reason to think that Davidson can't adopt such a line. One prominent lacuna in Davidson's account of rationalizing explanations is the problem of wayward or deviant causal chains. Consider the case of a mountain climber supporting his partner with a rope. He believes that this puts himself in a dangerous position and that loosening his hold on the rope would relieve him of the weight and the danger. These thoughts so unnerve the climber as to cause him to loosen his hold. Thus the relevant propositional attitudes, though a cause of the letting go of the rope, didn't cause the action in the appropriate way. Davidson writes that[14]

> Beliefs and desires that would rationalize an action if they caused it in the *right* way — through a course of practical reasoning — may cause it in other ways.

A coordinated belief-desire complex will cause a given action in the appropriate way just in case it was in virtue of a piece of practical reasoning. But that's just to say that content must be causally relevant in my sense. Davidson will never provide an adequate resolution to the problem of wayward or deviant causal chains without first addressing the metaphysical question — without first demonstrating how content can meet the necessary condition on causal relevance as specified in (3).[15] In fact it's not clear to me that one can even formulate the problem without at least tacitly presupposing this notion of causal relevance. The problem simply is that the reason for an action isn't causing it in the "right" way. What can appropriateness consist in other than the causal relevance of content?

It's clear that R^* must be a modal relation capable of itself supporting counterfactuals and subjunctive conditionals. It is sufficient that statements like:

(11) $\boxed{c} R^*(a, b)$

should come out true. Davidson thus can't use weak supervenience since weak supervenience only holds within a given world w. In adopting modal R^*, however, one doesn't give up token identity since it

will be trivially implied. At any world w $R =_t p$, for some physical event p. Thus Davidson's picture of the relationship of the mental to the physical, while failing to be a sufficient one, may well turn out to be a necessary one.

Would strong supervenience be sufficient to preserve the modal force of our vernacular belief-desire explanations? Consider then, property supervenience.[16] The mental will strongly supervene on the physical if and only if no two possible events instantiate the same physical properties while differing with respect to some mental property. Let P be the maximal physical property of a given agent, and let M be some mental property. If M strongly supervenes on P, then P will be causally sufficient for M. We thus have:

$$(12) \qquad \boxed{c}(\forall \mathbf{x})\,(\mathbf{P}_1(\mathbf{x}) \rightarrow \mathbf{R}(\mathbf{x}))$$

and

$$(13) \qquad \boxed{c}(\forall \mathbf{x})\,(\mathbf{P}_2(\mathbf{x}) \rightarrow \mathbf{A}(\mathbf{x}))$$

Where $\mathbf{x}$ ranges over individual, $\mathbf{R}$ is the predicate "is a primary reason R", $\mathbf{A}$ "is the intentional action A", and $\mathbf{P}_1$, $\mathbf{P}_2$ are predicates expressing the maximal physical property of the given individual. P, however, fails to be a necessary condition for M because of its strength, P being the *maximal* physical property of the agent. We need a weaker supervenience base. A given property A is weaker than property B if and only if whenever B exemplified A is, but not conversely. A property P' is a minimal physical base property just in case there is no weaker property which is causally sufficient for M. A supervenience thesis with P_1 as the supervenience base will thus provide causally necessary and sufficient conditions for M.

Kim, in 'Psyschophysical Supervenience as Mind-Body Theory', has shown that strong supervenience is compatible with Fodor's multiple realizability argument.[17] Multiple Realizability is the idea that a given macro-property can be realized by various micro-structures. Strong supervenience will allow for possibly infinite alternative minimal physical bases for a given mental property. Thus we have:

$$(14) \qquad \boxed{c}M \leftrightarrow \overset{\infty}{\underset{i=1}{V}} P.$$

There may of course be some relatively simple higher order physical property P^* corresponding to $V_{i=1}^{\infty} P$. Switching to talk of events we have:

(15) $\boxed{c}(\forall \mathbf{x})\,(\mathbf{R}(\mathbf{x}) \leftrightarrow \mathbf{P}_1^*(\mathbf{x}))$

and

(16) $\boxed{c}(\forall \mathbf{x})\,(\mathbf{A}(\mathbf{x}) \leftrightarrow \mathbf{P}_2^*(\mathbf{x}))$

where $\mathbf{x}$ ranges over events and $\mathbf{P}_1^*$, $\mathbf{P}_2^*$ are predicates expressing the appropriate higher order physical properties. From (15), (16) and

(17) $\boxed{c}(\forall \mathbf{x})\,(-\mathbf{P}^*(\mathbf{x}) \rightarrow -\mathbf{P}^*(\mathbf{x}))$[18]

we can derive counterfactuals like (1).

Despite the fact that (14) gives us necessary coextension between the supervenient mental properties and the base physical properties, it's not *prima facie* clear that this will insure the existence of psycho-physical bridge laws. As Fodor has pointed out, Boolean operations on the natural kinds of a given science may fail to preserve nomological force.[19] But Fodor never really provides an argument for this, and the suggestion may well turn out to be false. Even if it does pan out, even if Boolean operations really do fail to preserve nomological force, Davidson still might not be happy with this. Even though psycho-physical laws fail, multiple realizability is a much weaker reason for denying such laws than the Anomaly of the Mental. Davidson argues that psycho-physical laws fail because there's a categorical difference between mental and physical concepts. The former have rational conditions of application while the latter have non-rational conditions of application. Remember, a minimal physical base will be causally sufficient for some mental event. Strong supervenience effectively blurs the *categorical* difference between the mental and the physical, resulting, as it does, in a form of reductionism. With mutilple realizability you fix a token of a mental event and ask what various physical types it must be in order to exemplify the given mental type. Davidson, on the other hand, wants to fix the physical state of the agent and ask what various mental types he could exemplify. If Anomaly holds, no physical property will ever be causally sufficient for a mental property's instantiation.

Strong supervenience, while sufficient for insuring the modal force of our folk psychological explanations, seems to be too strong for Davidson's purposes. Perhaps there's some other candidate for modal R^*. The problem is that if R^* is a modal relation of the appropriate sort, we'll get necessary coextension. It's not clear to me how to clear out some kind of metaphysical middle ground — perhaps if one could find some restriction on the relevant possible worlds in evaluating counterfactuals through the use of the appropriate *ceteris paribus* clauses. All that would be required would be for the supervenience relation to hold in all accessible worlds. But I have no idea how to flesh out this suggestion since I don't know how to provide a non-ad hoc restriction for the *ceteris paribus* clause. I don't have an argument, but I suspect the relation required by Davidson may turn out to be metaphysically incoherent. Davidson's understanding of supervenience is based on a confusion — a confusion which made the supervenience relation so initially attractive. It was thought that supervenience could give one a counterfactual dependency relation without any corresponding reduction. Unfortunately, weak supervenience won't support the appropriate counterfactuals and strong supervenience gives way to type-type connections.

STRICT PHYSICALIST CAUSATION

Physicalism, which has served as the basic ontological framework as well as a methodological constraint in reaction to dualism for most of the recent work in the philosophy of mind, has very naturally influenced and constrained people's thoughts concerning psycho-physical causation. The general strategy has been to somehow construe psychophysical causation as a species of physical-physical causation. I want to suggest that we reexamine the kind of constraint that our adherence to strict physicalist causation has put on providing an account of the causal efficacy of propositional attitudes. I believe that such a strict physicalist notion of causation has had its philosophical costs for Davidson's account —it is precisely this which makes his notion of the mental epiphenomenal. I think that more flexibility on this count may give content theorists more room to manoeuver and may open some interesting philosophical alternatives.

So what exactly is strict physicalist causation? Intuitively, one can think of it as the doctrine that physical theory determines the range of causal possibility. One can see this most clearly in Davidson's Principle of the Nomological Character of Causality. Recall that Davidson holds the neo-Humean view that all causal relations must be nomologically subsumed by strict laws. Thus the complete specification of these strict laws will also completely specify the range of causal possibility, i.e., it will specify whatever is causally possible. Unfortunately it is unclear what exactly Davidson meant by restricting the nomic subsumption of singular causal statements to strict laws.[20] Despite the exegetical difficulties connected with this, Davidson explicitly requires that any system of strict laws have the properties of closure and comprehensiveness. Presumably these restrictions are there to insure that Davidson has completely captured all species of causal relations. None of this so far is what is causing problems. Rather, it is Davidson's intuitive and admittedly plausible claim that such strict laws are physical laws. Despite the prima facie plausibility and attractiveness of this move, that strict laws are only physical laws isn't the only logically possibility. It may turn out that there is some other system of laws that satisfy the closure and comprehensiveness constraint. Davidson's claim, however, goes unargued for. Davidson's account has the following property: since physical theory determines the range of causal possibility, and since the lawful descriptions instantiating these strict laws are presumably the appropriate micro-physical descriptions, all causal relations must satisfy the micro-physical supervenience condition — that all causal relations strongly supervene on the appropriate micro-physical states. The micro-physical supervenience condition is what makes strict physicalist causation *strict* physicalist causation. Rejection of the supervenience claim is tantamount to a rejection of strict physicalist causation.

It is clear that on Davidson's account, the semantic content of propositional attitudes just can't be causally relevant. Given his strict physicalist notion of causation, in order for content to be causally relevant, it would have to strongly supervene on physical properties. But given the Anomaly of the Mental which states that there is a *categorical* difference between mental and physical concepts (since mental concepts have rational conditions of application while physical concepts have non-rational conditions of applications) such a super-

venience claim would be repugnant. I'd like to suggest that a rejection of micro-physical supervenience with respect to causation might turn out to be a fruitful direction of research. I suspect that such a line will be the only way to provide an autonomist account of content.

Providing an account of physicalist causation which rejects the micro-physical supervenience condition is not an easy matter however. One immediately faces the following dilemma. Maintaining the strong supervenience condition is problematic for the autonomist as we have seen. However, if one substitutes weak supervenience for strong, physical theory won't determine the range of causal possibility, and it would seem that we no longer have an interesting *physicalist* notion of causation.

Narrow is the gate . . . .[21]

NOTES

[1] Cf. Jaegwon Kim: 1985, 'Concepts of Supervenience', *Philosophy and Phenomenological Research* 45, pp. 153—177. Weak supervenience is defined as follows:

$$\mathbf{A} \text{ weakly supervenes on } \mathbf{B} \Leftrightarrow \Box(\forall \mathbf{F} \in \mathbf{A})[(\forall \mathbf{x})(\mathbf{F}(\mathbf{x}) \rightarrow (\exists \mathbf{G} \in \mathbf{B})(\mathbf{B}(\mathbf{x}) \wedge (\forall \mathbf{y})(\mathbf{G}(\mathbf{y}) \rightarrow \mathbf{F}(\mathbf{y})))]$$

where **A** and **B** are sets of properties. Kim leaves the interpretation of the modality unspecified since different modalities may be appropriate for different contexts. Strong supervenience differs from weak in that it requires an additional modal operator:

$$\mathbf{A} \text{ strongly supervenes on } \mathbf{B} \Leftrightarrow \Box(\forall \mathbf{F} \in \mathbf{A})[(\forall \mathbf{x})(\mathbf{F}(\mathbf{x}) \rightarrow (\exists \mathbf{G} \in \mathbf{B})(\mathbf{B}(\mathbf{x}) \wedge \Box(\forall \mathbf{y})(\mathbf{G}(\mathbf{y}) \rightarrow \mathbf{F}(\mathbf{y})))].$$

[2] Cf. David Lewis' discussion of epiphenomena in: 1973, 'Causation', *Journal of Philosophy* 70, pp. 556—567.

[3] For example, similar problems should arise for Tyler Burge's account of the explanatory role of content in: 1986, 'Individualism and Psychology', *The Philosophical Review*, pp. 3—45.

[4] In 1982, *Essays on Actions and Events* (Clarendon Press, Oxford), pp. 3—19.

[5] Cf. 'The Explanatory Role of Content', forthcoming.

[6] Explanation is an epistemic notion. Explanations are given to engender understanding. When we need an explanation we are in some sort of epistemic predicament (what Sylvain Bromberger calls a "*p*-predicament" or "*b*-predicament" — cf. 1965. 'An Approach to Explanation', *Analytic Philosophy* (Basil Blackwell, Oxford)). A given explanation will be adequate, then, only if it resolves the tension inherent in such an epistemic situation. It would be interesting to more clearly specify the epistemological value of counterfactuals, i.e., their role in resolving our epistemic predicament.

[7] Cf. Ernest Lepore and Brian McLaughlin: 1985, 'Actions, Reasons, Causes, and Intentions', and Michael Bratman: 1985, 'Davidson's Theory of Intention', *Actions and Events* (Basil Blackwell, New York), pp. 3—13 and 14—28 respectively.

[8] Davidson writes: "If someone performs an action of type *A* with the intention of performing an action of type *B*, then he must have a pro-attitude towards actions of

type B . . . and a belief that in performing an action of type A will be . . . performing an action of type B . . . . The expression of the belief and desire entail that actions of type A are, or probably will be, good . . . . The descriptions of the action provided by the phrase substituted for 'A' gives the description under which the desire and belief rationalize the action."

In 1982, 'Intending', *Essays on Actions and Events*, pp. 86—87.

[9] *Ibid.*, pp. 83—102.

[10] In *Essays on Actions and Events*, pp. 207—225. Cf. also Brian McLaughlin: 1985, 'Anomalous Monism and the Irreducibility of the Mental', *Actions and Events*, pp. 331—367.

[11] Cf. Carl G. Hempel: 1965, *Aspects of Scientific Explanation* (The Free Press, New York).

[12] 1980, 'Oblique Causation and Reasons for Action', *Synthese* 43, pp. 351—367.

[13] The following has been inspired by Loewer and Lepore's criticism of Dual Aspect Semantics ('Dual Aspect Semantics', forthcoming) and Dretske's recent reductionist account of content ('Misrepresentation', 'The Explanatory Role of Content', and 'Explaining Behavior', forthcoming).

[14] 1982, 'Freedom to Act', *Essays on Actions and Events*, p. 79.

[15] I owe this point to Barry Loewer.

[16] Cf. Jaegwon Kim: 1985, 'Concepts of Supervenience', *op. cit.*, 1979, 'Causality, Identity, and Supervenience', *Midwest Studies* 4, pp. 31—49, and 1984, 'Epiphenomenal and Supervenient Causation', *Midwest Studies* 9, pp 257—270.

[17] 1982, *Brain and Cognitive Theory*. Cf. also Jerry Fodor, 1975, chapter 1 of *The Language of Thought* (Harvard University Press, Cambridge).

[18] This is, of course, question begging with respect to the Problem of Spurious Overdetermination, but is, nevertheless, sufficient for expository purposes. A more satisfactory account would take greater care in specifying the appropriate *ceteris paribus* clauses for Dretske's counterfactual condition.

Although having the content of the agent's propositional attitudes strongly supervene on the physical states and events of the agent's body is sufficient to capture Dretske's counterfactual condition, it is important to emphasize that securing the appropriate counterfactuals and subjunctive conditionals is only a *necessary* condition on causal relevance. It's *not* a sufficient condition. It isn't enough to establish some kind of counterfactual covariation between the intentional properties of a contentful mental event and its causally efficacious micro-physical properties in order to insure the causal relevance of the mental event's semantic content. If there's no causal chain between a primary reason R and an intentional action A independent of the causal chain between physical events p_1 and p_2, then what is to keep the causal powers of R from being pre-empted by p_1? As Kim has argued ('Causality, Identity, and Supervenience'), strong supervenience may itself be compatible with a weak form of epiphenomenalism since the supervenience base is doing all the work in psycho-physical causation.

It's not clear that Kim's weak epiphenomenalism is a stable position. One may be able to deploy the Problem of Pre-Emption in a general argument to the effect that Kim's position collapses into a form of eliminativism.

[19] Cf. *The Language of Thought*.

[20] Cf. Brian McLaughlin's 'Anomalous Monism and the Irreducibility of the Mental'.

[21] In fact, Barry Loewer suggests that the gate is closed. Nevertheless, my principal point remains good — that if you want an autonomist notion of content that's not epiphenomenal, you had better reject strict physicalist causation. For instance, one could account for the causal relevance of content in terms of Lewis' counterfactual analysis of causation.

In some unpublished notes Jaegwon Kim has independently developed a similar argument against Davidson. Kim's main criticism, like mine, is that while Davidson can

get the appropriate physical counterfactuals, he can't get counterfactuals in terms of the relevant intentional descriptions. Kim develops an interesting ancillary argument using the notion of "causal informativeness". A given description is causally informative if it is similar to or in the "vicinity" of descriptions in terms of which the relevant subsumptive nomic generalization is stated. A singular causal statement will be explanatory if it is given in terms of causally informative descriptions. Thus the following singular causal statement:

(A) The event reported on page 1 of The Times caused the event reported on page 3 of The Herald.

fails to be explanatory since it fails to be causally informative. Kim's argument takes the form of a challenge — given that there are no type-type connections between mental and physical descriptions, how is it that rationalizing explanations are any different from the causally *un*informative singular causal statement (A)?

I would like to thank Paul Boghosian, Don Demetriades, Jaegwon Kim, Barry Loewer, Dion Scott-Kakures, Brian McLaughlin, Sigrun Svavarsdottir, and Bill Taschеck for helpful criticism and comments. And special thanks to A B Carter and Gary Ebbs whose healthy philosophical scepticism was *sine qua non*.

REFERENCES

Bratman, M.: 1985, 'Davidson's Theory of Intention', *Action and Events* (Basil Blackwell, New York), pp. 14—28.
Bromberger, S.: 1965, 'An Approach to Explanation', *Analytic Philosophy* (Basil Blackwell, Oxford).
Burge, T.: 1985, 'Individualism and Psychology', *The Philosophical Review*, pp. 3—45.
Davidson, D.: 1982, *Essays on Actions and Events* (Clarendon Press, Oxford).
Dretske, F. I.: 'Misrepresentation', forthcoming.
Dretske, F. I.: 'The Explanatory Role of Content', forthcoming.
Dretske, F. I.: 'Explaining Behavior', forthcoming.
Fodor, J.: 1975, *The Language of Thought* (Harvard University Press, Cambridge).
Hempel, C. G.: 1965, *Aspects of Scientific Explanation* (The Free Press, New York).
Kim, J.: 1979, 'Causality, Identity, and Supervenience in the Mind-Body Problem', *Midwest Studies in Philosophy* 4, pp. 31—49.
Kim, J.: 1982, 'Psychophysical Supervenience as Mind-Body Theory', *Brain and Cognitive Theory*.
Kim, J.: 1984, 'Epiphenomenal and Supervenient Causation', *Midwest Studies in Philosophy* 9, pp. 257—270.
Kim, J.: 1985, 'Concepts of Supervenience', *Philosophy and Phenomenological Research* 45, pp. 153—177.
Lepore, E. and McLaughlin, B.: 1985, 'Actions, Reasons, Causes, and Intentions', *Actions and Events* (Basil Blackwell, New York), pp. 3—13.
Lewis, D.: 1973, 'Causation', *Journal of Philosophy* 70, pp. 556—567.
Loewer, B. and Lepore, E.: 'Dual Aspect Semantics', forthcoming.
McLaughlin, B.: 1985, 'Anomalous Monism and the Irreducibility of the Mental', *Actions and Events* (Basil Blackwell, New York), pp. 331—367.
Stoutland, F.: 'Oblique Causation and Reasons for Action', *Synthese* 43, pp. 351—367.

Department of Philosophy,
University of Michigan,
Ann Arbor, MI 49109,
U.S.A.