

*Supervenience and
mind*

SELECTED PHILOSOPHICAL ESSAYS

Jaegwon Kim



CAMBRIDGE
UNIVERSITY PRESS

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
40 West 20th Street, New York, NY 10011-4211, USA
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

© Cambridge University Press 1993

First published 1993

Reprinted 1995

Library of Congress Cataloging-in-Publication Data is available.

A catalogue record for this book is available from the British Library.

ISBN 0-521-43394-0 hardback

ISBN 0-521-43996-5 paperback

Transferred to digital printing 2002

Psychophysical laws

I

The question whether there are, or can be, psychological laws is one of considerable interest. If it can be shown that there can be no such laws, a nomothetic science of psychology will have been shown to be impossible. The qualifier 'nomothetic' is redundant: science is supposed to be nomothetic. Discovery, or at least pursuit, of laws is thought to be constitutive of the very nature of science so that where there are no laws there can be no science, and where we have reason to believe there are none we have no business pretending to be doing science.

At least in one clear sense, therefore, the absence of psychological laws entails the impossibility of psychology as a science. This need not be taken to mean that there can be no scientists, called 'psychologists' or 'cognitive scientists', who study psychological topics and write useful tracts about them. It is to say that whatever else they may be doing that is useful and worthwhile, they will not be producing *psychological theories*, comprehensive and integrated systems of precise general laws, couched in a characteristic theoretical vocabulary, on the basis of which mental phenomena could be explained and predicted. If such theory-based explanatory and predictive activities are what we suppose psychologists *qua* psychologists to be engaged in, recognition of the impossibility of psychological laws would force us to reconsider the nature of psychology as an intellectual enterprise. In what follows we shall touch on this general issue, but our main topic here is the question of the possibility of laws about psychological phenomena.

It is no surprise, then, that Donald Davidson, who has vigorously ar-

I am indebted to Akeel Bilgrami, Reinaldo Elugardo, Fred Feldman, Adam Morton, Bruce Russell, Nicholas White, and the members of my seminars in philosophy of mind at Michigan in 1979 and 1982. Brian McLaughlin gave an interesting set of comments on this paper when it was presented at the Davidson Conference; however, they have not been taken into account in preparing the present draft.

gued against the possibility of psychological laws, titled one of his papers on this topic 'Psychology as Philosophy'.¹ The intended contrast of course is with 'psychology as a science', an unattainable goal if his striking arguments are sound. In advocating the lawlessness of the mental he joins a small but influential group of philosophers who have taken a dim view of the scientific prospects of psychology. Norman Malcolm, for example, has produced a set of arguments, inspired by broadly Wittgensteinian considerations, against scientific psychology.² There are also Quine's disdainful strictures on Brentano's 'science of intention'.³ In this paper, however, we shall be concerned exclusively with Davidson's arguments contained in a series of three papers, 'Mental Events',⁴ 'Psychology as Philosophy', and 'The Material Mind',⁵ focusing especially on the first of these.

There are reasons for taking Davidson's arguments seriously and trying to get clear about them. The arguments are interesting and challenging, and have fascinated those interested in philosophy of mind; however, there is little agreement as to exactly how they are supposed to work.⁶ Many philosophers have an opinion about how successful these arguments are (the published verdicts have been almost uniformly negative thus far), but most appear to feel uncertain about the accuracy of their interpretations, or think that the interpretations fail to make the arguments sufficiently interesting or plausible. Above all almost everyone seems to find Davidson's arguments extremely opaque; it is not difficult to discern the general drift of his thinking or pick out the basic considerations motivating the arguments; however, delineating their structure pre-

1 In *Philosophy of Psychology*, ed. S. C. Brown (Harper & Row, New York, 1974). Reprinted in D. Davidson, *Essays on Actions and Events* (Clarendon Press, Oxford, 1980).

2 *Memory and Mind* (Cornell University Press, Ithaca, N.Y., 1977). See also Bruce Goldberg, 'The Correspondence Hypothesis', *Philosophical Review*, 77 (1968), pp. 438–54.

3 W. V. Quine, *Word and Object* (The Technology Press of M.I.T., Cambridge, Mass., 1960).

4 In *Experience and Theory*, ed. Lawrence Foster and J. W. Swanson (University of Massachusetts Press, Amherst, 1970). Reprinted in Davidson, *Essays on Actions and Events*.

5 In *Logic, Methodology, and the Philosophy of Science*, vol. 4, ed. P. Suppes (North-Holland, Amsterdam, 1973). Reprinted in Davidson, *Essays on Actions and Events*.

6 The following, I believe, is a representative list of published discussions of Davidson's arguments (I am not including those that primarily focus on 'anomalous monism'): C. Z. Elgin, 'Indeterminacy, Underdetermination, and the Anomalous Monism', *Synthese*, 45 (1980), pp. 233–55; William Lycan, 'Psychological Laws', *Philosophical Topics*, 12 (1981), pp. 9–38; Ted Honderich, 'Psychophysical Lawlike Connections and Their Problem', *Inquiry*, 24 (1981), pp. 277–303; Brian Loar, *Mind and Meaning* (Cambridge University Press, Cambridge, 1981), pp. 20–5; Robert Van Gulick, 'Rationality and the Anomalous Nature of the Mental', *Philosophy Research Archives*, 1983; William Larry Stanton, 'Supervenience and Psychophysical Law in Anomalous Monism', *Pacific Philosophical Quarterly*, 64 (1983), pp. 72–9.

cisely enough for effective evaluation and criticism is another matter. In this paper I propose a way of looking at what I take to be Davidson's principal argument against nomological psychology. The suggested interpretation is based on a simple leading idea, and will help us piece together a coherent picture of Davidson's overall views of the mental and relate it to a wider context. My aim here is essentially to interpret and expound, not to evaluate or criticize. But obviously I am embarking on this project because I think the argument to be extracted from Davidson is plausible, at least at first blush, and philosophically important. As I hope will become clear, Davidson's argument has far-reaching implications regarding some basic issues about the nature of mind, such as mental autonomy, the possibility of free agency, and the status of commonsense explanations of human actions, and points to a conception of the mental that I find both intriguing and appealing.

Davidson's apparent strategy in 'Mental Events' is, first, to establish the following lemma:

*Psychophysical Anomalism:*⁷ There are no psychophysical laws, that is, laws connecting mental and physical phenomena. In fact, there *cannot* be such laws,

and then use it to argue for the desired general thesis of psychological anomaly:

Anomalism of the Mental: 'There are no strict deterministic laws on the basis of which mental events can be predicted and explained.'⁸

The bulk of 'Mental Events' and 'Psychology as Philosophy' is devoted to establishing Psychophysical Anomalism, and much of the interest generated by these papers has been focused on Davidson's arguments for this thesis. In contrast the move from Psychophysical Anomalism to the full Anomalism of the Mental is made rather quickly and abruptly, within one short paragraph in 'Mental Events'; I shall make some suggestions about how this transition can be understood, but for the moment, and for much of this paper, we shall follow Davidson in concentrating on arguments for Psychophysical Anomalism.

7 The term 'psychophysical anomalism' is not Davidson's.

8 'Mental Events', p. 208 (page references to this article are to its reprinted version in Davidson, *Essays on Actions and Events*).

Davidson's conception of the psychological is based on *intentionality*. Expressions we use in attributing *propositional attitudes*, such as 'believe', 'fear', 'hope', and 'regret', are taken to constitute the basic psychological vocabulary; psychological laws then would be laws stated in terms of these intentional psychological expressions. Two questions may be raised about this way of understanding the psychological: first, whether it is broad enough to cover 'phenomenal states' or 'qualia', like pains and after-images, and second, whether it applies to the terms of trade of 'scientific psychology' or 'cognitive science' as it is practiced nowadays. These are large questions and cannot be taken up here; the second raises an issue about the relationship between 'commonsense psychology' and systematic psychology, a topic of much current interest,⁹ and I shall make some remarks relevant to it below. In any event, the conception of the psychological as intentional does capture a large core of our commonsense psychological vocabulary, and a successful argument for the impossibility of psychological laws on this conception of the psychological would be of great interest and importance.¹⁰ It would show, for example, that familiar explanations of actions in terms of an agent's beliefs and desires could not be nomological explanations backed by laws about beliefs, desires, and the like, as claimed by some writers (e.g., Carl Hempel¹¹). And it would imply a significant general conclusion: law-based systematic psychology, if such a thing is possible, would have to make a radical break with the framework of our vernacular psychological idioms and truisms, which forms the basis of our shared ability to describe and make sense of our own motives and actions as well as those of our fellow humans, and without which communal human life would be unthinkable.

The initial impression one is likely to get from Davidson's discussion of Psychophysical Anomalism is something like this: we are first offered a long list of features that characterize the mental but not the physical and, conversely, features of the physical not shared by the mental. For example, the mental is intentional and rational but the physical is neither;

9 For a sustained recent treatment see Stephen P. Stich, *From Folk Psychology to Cognitive Science* (The M.I.T. Press, Cambridge, Mass., 1983). See also Adam Morton, *Frames of Mind* (Oxford University Press, Oxford, 1980).

10 Davidson explicitly limits his arguments to intentional mental states, e.g., 'Comments and Replies' following 'Psychology as Philosophy' in *Essays on Actions and Events*, p. 240.

11 In the title essay of *Aspects of Scientific Explanation* (The Free Press, New York, 1965).

physical laws are 'homonomic' but what mental generalizations that there are are 'heteronomic'; combining mental and physical terms in a single statement is like mixing 'grue' with 'emerald'; Quinean indeterminacy besets the mental but nothing analogous obtains for the physical; and so on. We are then tempted to ask: does Davidson expect us to infer from these dissimilarities and divergences that there can be no laws connecting the two systems? But how can he? No simple list of differences between the two domains will have any tendency to show that no laws can connect them. When two arbitrary domains are considered, there is no a priori obvious reason to think there are lawful connections between them; nor need there be any obvious reason to think there are none. We would of course expect that any argument designed to show that there are, or that there are not, correlation laws, will make use of some properties of the two particular domains involved. So differences between the mental and the physical must count; but noting them can only be a starting point. The substance of the argument must show why, given just *these* differences, there can be no correlation laws.¹²

To fix the general picture in mind, consider a domain U of objects and two sets, F and G, of properties. For example, think of U as a set of medium-sized material bodies, F as a set of colors, and G as a set of shapes. We may suppose that each object in U has exactly one color in F and one shape in G. Here we would not expect to find regular correlations between colors and shapes; an object of a given color could be of any shape, and vice versa. Thus, we would not expect true generalizations of the form:

(A) Every object in U with color C has shape S.

Or of the form:

(B) Every object in U with shape S has color C.

But this is not to say that, contrary to our justified expectations, we may not in fact find, say:

12 After reviewing the differences noted by Davidson between the mental and the physical, Honderich writes: 'Still, we are not given a reason for thinking that [Psychophysical Anomalism] follows from the description of the two domains. As others have asked, what reason is there for thinking that an item which falls in one domain, and whose description then depends on X, cannot be in a lawlike connection with an item in the other domain, whose description then depends on Y? There is no general truth to the effect that there cannot be lawlike connection between items whose descriptions have different necessary connections. . . . Davidson remarks that his argument is no proof. It must also be said, I think, that his argument is at least crucially incomplete' ('Psycho-

(C) Every red object in U is round.

If this should happen, though, we would surely think it was pure luck, the result of a fortuitous choice of U. Given what we know about colors and shapes, we would not take the truth of (C) as indicating a *lawlike* connection between being red and being round; the truth of (C) is a coincidence, not a matter of law. We are especially unlikely to take it as lawlike if it is the single isolated correlation between colors and shapes; if it were a law we would expect it to be part of a broader system of color–shape correlations.

Turning to the matter at hand, consider the domain to be the set of persons, and F and G to be, respectively, the set of psychological properties and the set of physical properties. Davidson's point is that even if we should find a true generalization of the form:

(D) All persons with mental property M have physical property P,

we will not, and should not, consider this a *law*. What then is a law? Davidson follows the standard philosophical usage: a law is distinguished from a 'mere generalization' by these two marks: (1) it can support counterfactuals and subjunctives, and (2) it is confirmable by observation of instances. Our (C) above, about all red things being round, meets neither of these criteria; it fails to back a counterfactual such as 'If bananas were red, they would be round', and the only way it could be confirmed is by an exhaustive examination of all objects in the domain, there being no instance-to-instance accretion of positive confirmation.

It will be important to keep in mind the crucial role that considerations of lawlikeness must play in Davidson's central argument; for the argument is designed to show, not that there can be no true psychophysical generalizations of the form (D), but that there can be no psychophysical laws. Davidson is quite explicit on this point.¹³ And for good reason; brief reflection will show why this strategy is the only possible one: whether any generalization of the form (D) is true is a contingent empirical matter that can be known only through tedious observation, if at all. No armchair philosophical argument can insure that some statements of this form, by sheer luck or coincidence, will not turn out to be true; it surely cannot do this any more than it can show that a generalization like (C), that all

physical Lawlike Connections and Their Problem', pp. 292–3). This reaction is typical and understandable. What I intend to do is to help complete Davidson's argument.

¹³ 'Mental Events', p. 216.

red things are round, is true, or that it is false. Moreover, whether something like (C) or (D) is true is of no philosophical interest; what is of interest is whether, if true, it would be a law. Davidson thinks we can show from the very idea of what it is to be psychological that no generalization of the form (D), whether true or false, can be lawlike. Its being lawlike is independent of its de facto truth or falsity, and hence can be established or refuted by a priori arguments. At least, that is Davidson's view.

These considerations suggest a clue to the structure of Davidson's argument: *the argument works, to the extent that it does, only with respect to psychophysical laws, and it should fail, more or less obviously, if 'true psychophysical generalization' is substituted for 'psychophysical law' throughout the argument.* What needs to be done, therefore, is to identify the features of the mental and those of the physical that, while tolerating true psychophysical generalizations, are inimical to these generalizations being lawlike. And if this is to be done, the argument must consciously exploit the special characteristics of laws that set them apart from de facto generalizations.

III

The leading idea of Davidson's argument as I see it can be introduced through an analogy. Most of us remember being told by politicians or political analysts that a democratic nation cannot, on pain of damaging its own integrity as a democracy, enter into a genuine treaty relationship with a totalitarian state. We can also imagine something like this said of two religions: the systems are so alien to each other that no regularized and stable relationship between them is possible. The hidden argument here may be something like this: two systems of government or religion are so fundamentally opposed to each other in their basic commitments that a stable and principled relationship cannot be maintained between them on pain of compromising the integrity of one or both of the systems involved. It might be that a democratic state, if it is to honor its treaty obligations to a totalitarian state, must of necessity violate its own commitment to democratic principles. A weaker relationship could be tolerated, but treaty relations are too strict and binding, imposing on the participants obligations that weaker relations do not impose.

Whatever merits the foregoing might have, the structure of Davidson's argument, I believe, is similar: *the mental realm is characterized by certain essential features which would be seriously compromised if there were connections as strong as laws, with their modal and subjunctive force, linking it with the physical realm,*

which has its own distinctive essential features incompatible with those of the mental. These features of the mental are essential in that they are *constitutive* or *definitive* of the system of mental concepts; the mental realm cannot sustain their loss and still retain its identity as a mental system. Further, these features are *global* in the sense that they characterize the mental as a system, not primarily individual mental phenomena or concepts in isolation; and similarly for the essential features of the physical. The argument could be run the other way also: given its own commitment to certain constitutive principles not shared by the mental, the physical realm can no more readily tolerate nomological relationships with the mental, without endangering its identity as a physical system. Mere psychophysical generalizations, being weaker than laws, do no harm to either psychology or physics, but laws with their modal force would bring them too close together, leading to a clash of their incompatible natures.

This way of looking at Davidson's argument explains exactly why the argument is supposed to work for psychophysical laws but not for true psychophysical generalizations. True generalizations, unless they are law-like, are merely accidental and do not signify any deep or intimate relationship between the two realms (recall the case of colors and shapes). But laws are different: nomic connections are strong enough to *transmit*, or *transfer*, the constitutive properties of the physical to the mental, and vice versa, thereby damaging the integrity of the recipient system. Mere generalizations, even if true, do not have this power of transmitting features of one system to the other. We shall try to fill out this preliminary sketch of the argument by giving concrete meaning to this idea of laws 'transmitting' certain features across systems, but a hint of how this can be understood is contained in the observation that laws, in virtue of their modal force, can underwrite certain inferences that mere *de facto* generalizations cannot sanction.

The skeletal structure of what I take to be Davidson's principal argument can, therefore, be exhibited as follows:

The mental system has a certain essential characteristic X and the physical system a certain essential characteristic Y, where X and Y are mutually incompatible. Laws linking the two systems, if they exist, would 'transmit' these characteristics from one system to the other, leading to incoherence. Therefore, there can be no laws connecting the mental with the physical so long as the two systems are to retain their distinctive identities.

It is worth pointing out that the argument as sketched has a general interest going beyond its application to the psychophysical case; if appropriate properties X and Y are identified, the argument would apply to any two

domains and help establish the conclusion that there could not be lawlike connections between them. In any event, the proposed line of interpretation explains, and is supported by, the following remarks by Davidson: 'If the case of supposed laws linking the mental and the physical is different, it can only be because to allow the possibility of laws would amount to changing the subject. By changing the subject I mean here: deciding not to accept the criterion of the mental in terms of the vocabulary of the propositional attitudes.'¹⁴ Davidson is saying that if there were psychophysical laws we would lose the mental ('change the subject') as characterized in terms of intentionality; such laws would compromise the essential intentionality of the mental.

Two things need to be done to flesh out the skeletal argument: (1) we need to identify one or more essential characteristics of the mental, and do the same for the physical, to play the role of X and Y, that is, to be transmitted, or be compromised, by the supposed laws between the mental with the physical, and (2) we must explain in what sense laws can 'transmit' these characteristics from one system to the other. Let us turn to the first task.

Davidson does not tell us in a general way what he means by 'intentional'; instead he simply tells us that the paradigmatic mental states he has in mind are *propositional attitudes*, that is, psychological states with *propositional content* typically expressed by that-clauses and gerunds (e.g., fearing that the pipes are frozen, being embarrassed about missing his appointment for the second time). What then are the crucial features of such states that can be used to fill out Davidson's argument? Consider the following remarks by Davidson:

Any effort at increasing the accuracy and power of a theory of behavior forces us to bring more and more of the whole system of the agent's beliefs and motives directly into account. But in inferring this system from the evidence, we necessarily impose conditions of coherence, rationality, and consistency. These conditions have no echo in physical theory, which is why we can look for no more than rough correlations between psychological and physical phenomena.¹⁵

Just as we cannot intelligibly assign a length to any object unless a comprehensive theory holds of objects of that sort, we cannot intelligibly attribute any propositional attitude to an agent except within the framework of a viable theory of his beliefs, desires, intentions, and decisions.

There is no assigning beliefs to a person one by one on the basis of his verbal behavior, his choices, or other local signs no matter how plain and

¹⁴ *Ibid.*, p. 216.

¹⁵ 'Psychology as Philosophy', in *Essays on Actions and Events*, p. 231.

evident, for we make sense of particular beliefs only as they cohere with other beliefs, with preferences, with intentions, hopes, fears, expectations, and the rest. It is not merely, as with the measurement of length, that each case tests a theory and depends upon it, but that the content of a propositional attitude derives from its place in the pattern.¹⁶

These remarks vividly bring out Davidson's 'holism' of the mental: the mental is holistic in that the attribution of any single mental state to a person is strongly constrained by the requirement that the total system attributed to him of beliefs, desires, fears, hopes, and all the rest be *maximally coherent and rational*. This coherence or rationality maximization condition, on Davidson's view, is an essential feature of the intentional; without it we cannot make sense of ascription of contentful mental states. The holistic character of the mental, as embodied in the principle of rationality maximization, is constitutive of our conception of the mental as intentional; compromising this characteristic of the mental would be tantamount to 'changing the subject' – that is, as Davidson explains, abandoning the intentional conception of the mental. How does one maximize the coherence and rationality of a system of intentional states? This is an age-old issue of great importance to epistemology, moral philosophy, and philosophy of science, and we need not address it in a general way. What we need is a sense of what it is about. To begin, avoiding logical inconsistency and maximizing inductive rationality in one's belief system is obviously important; the internal coherence of the agent's system of preferences, e.g., that it satisfy the transitivity condition, is also a factor; we should also check whether the agent's decisions conform to his probabilities and preferences, and whether his feelings and emotions make sense in light of his wants and beliefs; and so on. Davidson's view is, to put it briefly and somewhat simplistically, that either the set of intentional states we attribute to a person satisfies certain minimal standards of rationality and coherence, or else there is no ground for attributing such a system to an agent; in fact, to consider an organism an *agent* is an expression of our willingness to consider it a rational psychological system, that is, to describe its behavior in terms appropriate for assessment in accordance with canons of rationality, and make sense of its decisions and actions as issuing in appropriate ways from its preferences and cognitions. We might add that the point of attributing intentional states to persons is to be able to formulate 'rationalizing explanations' of what they do, and that unless the system of intentional states so attributed is, in certain minimal ways, rational and coherent, no such

16 'Mental Events', p. 221.

explanations would be forthcoming. Davidson says that Quine's doctrine of translational indeterminacy is just another facet of this rationalistic holism of the mental; I shall make a few remarks later about how these two theses are related, but I believe we can construct an argument for Psychophysical Anomalism without an explicit reference to, or reliance on, the indeterminacy thesis.¹⁷

In point of being holistic, however, the mental is not unique; on Davidson's view, the physical, too, is holistic. Interdependence or seamlessness is common to both.¹⁸ The holism of the physical lies in the fact that the physical, too, is characterized by certain 'synthetic a priori laws' which are *constitutive of* our conception of the physical, and which make possible the formulation of precise physical laws. Among them are principles that make physical measurement possible, such as the transitivity of 'longer than' or 'earlier than'.¹⁹ Basic methodological rules governing theory construction and evidence, fundamental principles about space, time, and causality, and so on, may also qualify. Holism as such, therefore, is a side issue; what is crucial is the divergent constitutive principles from which the distinctive holism of each domain arises. As Davidson puts it, 'there are no strict psychophysical laws because of the disparate commitments of the mental and the physical schemes',²⁰ and 'there cannot be tight connections between the realms if each is to retain its allegiance to its proper source of evidence'.²¹

These two brief remarks by Davidson are especially revealing: the mental and physical are not able to 'keep allegiance' to their respective constitutive principles and at the same time enter into the kind of 'tight connection' signified by the presence of laws linking them. For the two sets of constitutive principles represent the 'disparate commitments' of the two systems, commitments they cannot disown if they are to preserve their identities. What we now need to understand is exactly how the presence of nomological links is inconsistent with each system's retaining its allegiance to its constitutive principles.

If rationality, therefore, is the essential characteristic of the mental in Davidson's argument, what is the essential feature of the physical that will clash with rationality? I believe we can simply take this as the absence of rationality as a constitutive element of the physical. As Davidson says in a passage already quoted, conditions of coherence, rationality, and consistency 'have no echo in physical theory'.

17 As Lycan emphasizes in 'Psychological Laws', p. 23. 18 'Mental Events', p. 222.
19 *Ibid.*, pp. 220–1. 20 *Ibid.*, p. 222. 21 *Ibid.*, p. 222.

IV

My suggestion is that we try to understand the crucial step in Davidson's argument in terms of the greater inferential strength of laws, compared with de facto generalizations, on account of their modal force. I shall now formulate two specific arguments based on this idea.

Suppose that, on available evidence, the attribution to a person of either of the two mental states, m_1 or m_2 , is warranted, and that the principle of rationality maximization enjoins the choice of m_1 over m_2 (we may suppose that the joint attribution of both states contravenes this principle). Suppose further that there are neural states, n_1 and n_2 , which are *nomologically coextensive* with m_1 and m_2 respectively; that is, we have laws affirming that as a matter of law, n_1 occurs to an organism at a time just in case m_1 occurs to it at that time; similarly for n_2 and m_2 . Now the neural states, n_1 and n_2 , being theoretical states of physical theory, have *conditions of attribution*, that is, conditions under which their attribution to an organism is warranted. Such conditions are probably very complex and in some sense holistic; they are probably difficult to articulate, and we are not assuming that they must be observationally accessible. What matters is only that the ascertaining of whether they hold in a given situation is regulated by the constitutive rules and principles of physical theory, not by those of the mental. To say that C_1 is an attribution condition for n_1 must be more than to affirm a mere de facto coincidence of C_1 with n_1 (or with the warranted attribution of n_1); it is to commit oneself to a statement with modal force, which for simplicity we may express as follows:

(1) Necessarily, if C_1 obtains, n_1 occurs.

We also have the psychophysical law:

(2) Necessarily, m_1 occurs if and only if n_1 occurs,

whence:

(3) Necessarily, if C_1 obtains, m_1 occurs.

In the same way we have:

(4) Necessarily, if C_2 obtains, m_2 occurs,

where C_2 is an attribution condition of neural state n_2 .

Consider the force of (3) and (4): they affirm that when a certain set of physical conditions holds, a specific mental state *necessarily* occurs, that

we *must* attribute to an organism this mental state if those conditions are observed to obtain for it. And this means that the rationality maximization principle as an essential constraint on the attribution of mental states is in danger of being preempted, or seriously compromised, for the determination of whether these physical attribution conditions obtain is not subject to the constraint of this principle. Statements (3) and (4) would permit us to attribute intentional mental states independently of the rationality maximization rule; at least, they would force this rule to share its jurisdiction over mental attributions. In this way, these mental states threaten to escape the jurisdiction of the ruling constitutive principle of the mental, thereby losing their 'allegiance to [their] proper source of evidence'. By becoming so intimately associated with C_1 and C_2 , which are under the jurisdiction of physical theory and its constitutive principles, they have in effect ceased to be mental states. For according to Davidson, being subject to the rule of rationality maximization is of the essence of intentional states; without this constraint the ascription of contentful intentional states would be unintelligible.

If something like this captures Davidson's argument, then we should not be able to run it without the assumption that the supposed psychophysical correlations are lawlike; this is the assumption (2) above. It is obvious that if the modality is removed from (2) we can no longer move from (1) to (3), although we could get the nonmodal analogue of (3) stating a de facto coincidence, 'if C_1 obtains, m_1 occurs'. But this is harmless; it exerts no pressure on the rationality maximization principle as a constraint on the attribution of m_1 . A de facto conditional like this cannot be taken as stating an attribution condition of m_1 , no matter how loosely we construe the notion of attribution condition.

Two points in this argument require further comments. The first concerns the assumption that mental states m_1 and m_2 have *coextensive* physical correlates. This assumption simplifies the argument and enables us to derive a salient and striking conclusion; but it can be weakened. Obvious further cases to consider would be, first, one where neural state n is only sufficient for mental state m and, second, one where n is only necessary for m . However, these do not exhaust all the possibilities, and it will be useful to consider this in a fully general setting. So let $L(m,n)$ be an arbitrary law linking m and n . If this law is properly to be thought of as 'linking' m and n , then the logical form of $L(m,n)$ must generate strong mutual constraints between the attribution of m and that of n . To assume m and n to be coextensive is to set these constraints at a maximum level; if n is only sufficient, or only necessary, for m , the constraint is weaker

but still quite strong. Now, the generalized argument for arbitrary $L(m,n)$ would be something like this: If $L(m,n)$ is to qualify as psychophysical law, the attribution of m to an organism must strongly constrain, and be strongly constrained by, the attribution of n to that organism, and to that extent the constitutive principles of one domain extend their regulative powers to the other domain, thereby infringing upon the latter's integrity and autonomy.

The second point concerns the modalities involved in the displayed statements (1)–(4); more specifically, a question can be raised whether the 'nomological modality' of (2) is the same as the modality involved in the statement of 'attribution condition' (1). This raises a host of complex issues which are best avoided here; a short and reasonable way to handle the point would be this: assume that the modality involved in (1) is that of unrestricted logical necessity, and that logical necessity entails nomological necessity. This would imply that the modality of (3) and (4) is at least as strong as the nomological modality of (2); the crucial step would be to argue that this is sufficient to make (3) and (4) a threat to the mentalistic identity of m_1 and m_2 . If the likes of (3) and (4) were to hold, that would generate a strong pressure to integrate these affected mental states into physical theory.

We now turn to another way of filling out our skeletal argument. Let p be the statement 'Ypsilanti is within 10 miles of Ann Arbor' and q the statement 'Ypsilanti is within 20 miles of Ann Arbor'. The rule of rationality maximization presumably requires that whenever we attribute to a person the belief that p we must also attribute to him the belief that q . This much deductive closure seems required of any system of beliefs. Consider the following counterfactual:

(5) If S were to believe p , S would also believe q .

This dependence is *grounded in* the principle of rationality maximization; in fact, this principle may sanction a more specific principle enjoining us to attribute to a person all obvious logical consequences of beliefs already attributed to him. That (5) obtains is an important fact about the concept of belief, and is explainable in terms of the essential features of belief as an intentional state, that is, in terms of considerations of rationality and coherence of intentional systems. Suppose now that believing p and believing q have nomological coextensions, B_1 and B_2 respectively, in physical theory. We construe this to mean, or imply, the following:

- (6) Necessarily, a person believes p if and only if he is in state B_1 .
- (7) Necessarily, a person believes q if and only if he is in state B_2 .

Inferences involving counterfactuals are tricky; however, we may assume that (5), (6), and (7) together yield:

(8) If S were in state B₁, he would also be in state B₂.

Now, (8) is a *purely physical* counterfactual stating a dependency relation between two physical states; it might state a lawful dependency relation between two neurophysiological states involving discharges of large groups of neurons, or something of the sort. The fact that (6) and (7), the supposed psychophysical laws, would enable us to 'read off' a physical law from a psychological law is not the heart of the argument. We get closer to it when we ask: *What could possibly ground or explain this physical dependency?*

What then would explain or ground (8)? There are three possibilities to consider:

(a) The dependency expressed by (8) is physically fundamental – it is a basic law of physical theory requiring no explanation. This is highly implausible: we would expect fundamental physical laws to connect physical states a good deal simpler than neural correlates of beliefs.

(b) Statement (8) is explainable in terms of more fundamental physical laws. In this case, the same physical laws would yield, via (8), (6), and (7), a physical explanation of why the psychological dependency relation (5) holds, and this means that the role of the rationality maximization principle as a ground for (5) has been preempted, and that the concept of belief has effectively been removed from the jurisdiction of this principle. But the concept of belief that is outside the domain of rationality is no longer an intentional concept – not a concept of belief at all.

(c) The dependency relation (8), though not regarded as a basic physical law, has no physical explanation. But then we can explain it psychologically in terms of (5) via (6) and (7), as it was originally derived. But this is absurd: to ground a purely physical dependency in considerations of rationality of belief would have to be taken as an intolerable intrusion on the closedness and comprehensiveness of physical theory. Thus, none of the possibilities makes sense, and we must reject the supposed laws such as (6) and (7).

This concludes my attempt to flesh out Davidson's idea that psychophysical laws would bring too close together two systems with their 'disparate commitments'. There are no doubt other, perhaps more plausible, ways of doing so; however, what has been done here, I think, goes some way toward making Davidson's arguments more concrete and more palpable, and in my view not altogether implausible.

One might ask why we could not show, by the same argument, that there could not be laws connecting, say, biological and physical phenomena. The answer is that biology and physics are both physical theories sharing the same fundamental constitutive principles; they are governed not by 'disparate commitments' but one uniform set. I think this would be Davidson's response.²² If, on the other hand, you believe in the uniqueness of 'vital phenomena' or 'entelechies', you could make up a Davidsonian argument to show the nomological irreducibility of the vital to the physical; your only problem would be to defend the relevant vitalistic premises.

There are some prominent considerations advanced by Davidson, especially in 'Mental Events', that have not been made use of in my interpretation. The distinction between 'homonomic' and 'heteronomic' laws is one example; another is his likening of psychophysical laws to the mixing of 'grue' and 'emerald'; I have already mentioned Davidson's approving references to translational indeterminacy. My view is that these do not, at least need not, play a crucial role in the argument. In 'Mental Events', the distinction between the two types of laws quickly leads into the discussion of synthetic a priori constitutive principles of physical theory, and this latter idea of course plays a role in my interpretation. I take the reference to 'grue'-like predicates as just a way of illustrating the incongruity that exists, in Davidson's eye, between mental and physical terms, an incongruity that, as we saw, is given a more precise meaning in terms of allegiance to disparate sets of constitutive principles.

V

I shall now briefly consider how Psychophysical Anomalism relates to the Anomalism of the Mental. In 'Mental Events', one gets a strong impression that Davidson intends to infer the latter from the former. The following is the crucial paragraph:

It is not plausible that mental concepts alone can provide [a comprehensive framework for the description and law-based prediction and explanation of events], simply because the mental does not . . . constitute a closed system. Too much happens to affect the mental that is not itself a systematic

22 Actually what Davidson says about this is noncommittal: 'I do not want to say that analogous remarks may not hold for some other sciences, for example biology. But I do not know how to show that the concepts of biology are nomologically irreducible to the concepts of physics. What sets apart certain psychological concepts – their intentionality – does not apply to the concepts of biology' (in 'Comments and Replies' following 'Psychology as Philosophy', in *Essays on Actions and Events*, p. 241).

part of the mental. But if we combine this observation with the conclusion that no psychophysical statement is, or can be built into, a strict law we have the principle of the Anomalism of the Mental: there are no strict laws on the basis of which we can predict and explain mental phenomena.²³

Davidson seems to be saying that we can infer the Anomalism of the Mental from the two premises: Psychophysical Anomalism and the statement that the mental, as distinguished from the physical, does not constitute a closed system. But how is the inference supposed to work?

I have no bright idea on interpreting this passage to yield a perspicuous and plausible argument. Instead I suggest another way of viewing the situation which, though possibly not Davidson's, is not altogether implausible and which seems to fit the large dialectic plan of 'Mental Events'. First, the Anomalism of the Mental can be thought of as being equivalent to the conjunction of Psychophysical Anomalism and the following thesis:

*Psychological Anomalism:*²⁴ There are no purely psychological laws, that is, laws connecting psychological events with other psychological events, which can be used to explain and predict these events.

If mental phenomena can be nomologically explained and predicted, then the required laws would have to be either psychophysical or purely psychological. Psychophysical Anomalism says laws of the first kind are not there; Psychological Anomalism says laws of the second kind are not there either. So there are no laws to explain and predict mental phenomena, and this is precisely the Anomalism of the Mental.

Thus, I see the Anomalism of the Mental simply as a conjunction of the two doctrines, Psychological Anomalism and Psychophysical Anomalism. This raises the question where Psychological Anomalism comes from. No readily identifiable argument for it can be found in 'Mental Events', although there is no question that Davidson is committed to it. Furthermore, there are passages in this paper that strongly suggest that the mental as an autonomous realm ought to have, or at least can have, its own laws. In particular, I have in mind Davidson's claim that the synthetic a priori constitutive principles of the physical domain are what makes 'homonomic' physical laws possible, and his explicit acknowledgement that the

²³ 'Mental Events', p. 224.

²⁴ The term 'psychological anomalism' is not Davidson's. Davidson is clearly committed to this thesis; for example, his argument for anomalous monism cannot go through unless it is assumed that there are no purely psychological laws; it isn't enough merely to assume there are no psychophysical laws.

mental domain, too, has its own characteristic a priori constitutive principles. He says, too, that the attribution of propositional attitudes presupposes as a necessary condition a 'viable *theory* . . . of beliefs, desires, intention, and decisions'.²⁵ What is a theory made up of, if not laws? But how can this be reconciled with Psychological Anomalism, or indeed with the Anomalism of the Mental?

I suggest the following line of reconciliation: on Davidson's account the mental can, and does, have its own 'laws'; for example, 'laws' of rational decision making. The crucial point, though, is that these are *normative* rather than *predictive* laws. When Psychophysical Anomalism and Psychological Anomalism deny the existence of laws about the mental, the meaning of 'law' involved is one that is appropriate to physical theory, namely the concept of law that permits the formulation of nomological predictions and explanations on the basis of precisely characterized and empirically identifiable initial and boundary conditions. It may be recalled that the Anomalism of the Mental only denies the existence of (in Davidson's own words) 'strict laws on the basis of which behavior can be explained and predicted'. Thus, the existence of nonpredictive normative laws or principles is consistent with the Anomalism of the Mental and Psychological Anomalism. But what do these normative laws look like? I already mentioned principles of decision making; rules of deductive and inductive inference, appropriately phrased, should also be among the prominent examples; there may be principles that govern the coherence of emotions, both among themselves and in relation to other propositional attitudes such as beliefs and desires. These are the norms and rules that guide actions and decisions, and form the basis of rational evaluations of our motives, cognitions, and emotions. And I think there is a sense in which these principles serve as an essential basis for a certain special way in which actions and decisions can be understood and made intelligible.²⁶ The view of psychology that emerges from Davidson is one of a broad interpretative endeavor directed at human action, to understand its 'meaning' rather than search for law-based causal explanations that are readily convertible into predictions; psychology is portrayed as a hermeneutic inquiry rather than a predictive science.

In order to appreciate Davidson's overall aims and strategies in 'Mental Events', it is useful to attend to his initial stage-setting. His announced aim, which he likens to Kant's attempt to reconcile human freedom with

25 'Mental Events', p. 221 (emphasis added).

26 This is developed in somewhat greater detail in my 'Self-Understanding and Rationalizing Explanations', *Philosophia Naturalis*, 21 (1984): 309–20.

natural necessity, is to show how psychological anomaly is compatible with determinism. How is it possible for the mental to escape the nomological net of physical theory? How can this happen when mental phenomena apparently enter into intimate causal transactions with physical phenomena? In order to formulate this problem, something like Psychological Anomalism has to be *presupposed*; psychological anomaly is part of Davidson's starting point in 'Mental Events' rather than a conclusion to be proved. In the second paragraph of 'Mental Events' he says, 'I start from the assumption that both the causal dependence, and the anomalousness, of mental events are undeniable facts'. Thus, three elements are needed to generate the initial 'Kantian tension': psychological anomaly, the causal dependence of the mental upon the physical, and physical determinism. The tension consists in our need to answer this question: how can the mental be anomalous (i.e., escape physical determinism) when it is causally dependent on the physical domain governed by strict deterministic laws? How can we protect the anomalousness, and the autonomy, of mind?

As I see it, Davidson's resolution consists in pointing out, first, that the tension arises because psychophysical causal dependence is erroneously thought to require the existence of psychophysical laws, and then showing that there in fact can be no such laws to threaten mental anomaly. His argument for the first point leads to his celebrated defense of 'anomalous monism', a version of the so-called 'token-identity' theory; but from the viewpoint of the overall aims of 'Mental Events', anomalous monism is a side issue. In any event, it is assumed in all this that psychophysical laws would make the mental reducible to the physical, effectively destroying its autonomous character. Thus, Psychophysical Anomalism is what safeguards Psychological Anomalism, by insulating the mental from the full impact of physical determinism. This is why arguments for Psychophysical Anomalism occupy center stage in 'Mental Events', and why, on the other hand, there are no arguments for Psychological Anomalism. The former is the substance of what has to be established to answer the principal question of 'Mental Events'; the latter only a presupposition of that question.

Has Davidson ever offered an independent argument for Psychological Anomalism? I believe he has; his discussion of the problem of empirically confirming Ramsey-style decision theory in 'Psychology as Philosophy' can usefully be viewed as just such an argument. But a detailed discussion of this argument is outside the scope of this paper.

In this concluding section I want to try to relate Davidson's views of the mental to a broader context. His initial Kantian tension can be re-described (by replacing psychophysical causal dependence with psychophysical laws) to yield an inconsistent triad: (1) psychological anomaly, (2) physical determinism, and (3) lawlike linkages between the psychological and the physical. Faced with this triad, Davidson rejects (3), and that is his Psychophysical Anomalism. And the ultimate goal of this move is to insure the autonomy of the mental and the possibility of free agency.²⁷ It is instructive, I think, to compare Davidson's move with Quine's: Quine, too, would accept (2) and reject (3), where (3) of course is understood to concern the psychological conceived as the intentional. In fact, his doctrine of translational indeterminacy can be taken as the denial of the claim that the intentional psychological *supervenes* on the physical; on Quine's view, the fixing of the totality of physical fact does not suffice to fix the intentional. If there were a pervasive network of laws linking the intentional with the physical, then the intentional would supervene on the physical.²⁸ Davidson and Quine, however, part company in their reaction to this failure of supervenience: while Davidson takes it as insuring the autonomy of the mental, Quine takes it as showing the illegitimacy of the mental, as witness his well-known disparaging remarks about Brentano: 'One may accept the Brentano thesis [of the irreducibility of intentional terms] either as showing the indispensability of intentional idioms

27 Davidson writes, at the very end of 'Mental Events' (p. 225): 'The anomalism of the mental is thus a necessary condition for viewing action as autonomous.' It is no accident that he begins and ends his paper with quotations from Kant.

28 See Quine's reply to Noam Chomsky in *Words and Objections*, ed. D. Davidson and J. Hintikka (Reidel, Dordrecht, 1969), esp. p. 303, where he says: 'Consider, from this realistic point of view, the totality of truths of nature, known and unknown, observable and unobservable, past and future. The point about indeterminacy of translation is that it withstands even all this truth, the whole truth about nature.' I am aware that in 'Mental Events', p. 214, Davidson explicitly endorses supervenience of the mental upon the physical, in spite of the nonexistence of psychophysical laws. To make Davidson consistent, however, this supervenience must be taken in a fairly weak sense falling well short of full dependence or determination. I am here using the term 'supervenience' in a stronger sense in which what supervenes is wholly fixed when the supervenience base is fixed. The distinction between 'weak' and 'strong' supervenience and related matters are developed in detail in my 'Concepts of Supervenience', Essay 4 of this volume; a simpler and somewhat sketchier account is included in my 'Psychophysical Supervenience as a Mind-Body Theory', *Cognition and Brain Theory*, 5 (1982), and 'Supervenience and Supervenient Causation', *Southern Journal of Philosophy*, 12, supplement (1984), pp. 45-56.

and the importance of the autonomous science of intention, or as showing the baselessness of intentional idioms and the emptiness of a science of intention. My attitude, unlike Brentano's, is the second.'²⁹ For Quine, reducibility to an extensional physical base is an essential mark of legitimacy. Davidson sees it as a threat to autonomy.

So there are two choices: the eliminativist physicalism of Quine and the dualism of Davidson. It undoubtedly will strike many readers as at best paradoxical to characterize Davidson as a dualist. I believe, however, that in spite of his anomalous monism, dualism in the form of a commitment to the mental as an autonomous domain is a nonnegotiable premise of Davidson's overall position in 'Mental Events'.

From this general perspective, we can also make sense of Davidson's somewhat cryptic remarks in 'Mental Events' linking Quine's thesis of translational indeterminacy with his Psychophysical Anomalism.³⁰ The essential function served by both doctrines is to pry apart the mental and the physical, and show the former to be irreducible, in a crucial way, to the latter. Where Davidson differs from Quine is in his attitude to this irreducibility. His attitude is strongly reminiscent of the dualism of Kant; it clearly is not Cartesian dualism – his anomalous monism is in effect the rejection of the interactionist dualism of the Cartesian variety.

One question remains: is there any reason for favoring this Kantian stance of mental autonomy over Quinean eliminativism? Alchemy and astrology are also irreducible to physical theory; we do not expect to find laws linking alchemical or astrological concepts with those of physics. But that hardly is any reason to champion an autonomous realm of alchemy or astrology! Here a Quinean response seems absolutely appropriate: so much the worse for alchemy and astrology! The irreducibility, nomological or conceptual, of these alleged inquiries to physical theory is conclusive evidence of the hollowness of their pretensions as serious theories of the world. Why should the case of the mental be different? This is a question of critical importance to the status of the mental in our scheme of things.

I think there is an answer, though this may not be Davidson's. The intentional psychological scheme – that is, the framework of belief, desire, and will – is one within which we deliberate about ends and means, and assess the rationality of actions and decisions. It is the framework that makes our normative and evaluative activities possible. No purely descriptive framework such as those of neurophysiology and physics, no

29 *Word and Object*, p. 221. 30 'Mental Events', p. 222.

matter how theoretically comprehensive and predictively powerful, can replace it. As long as we think of ourselves as reflective agents capable of deliberation and evaluation – that is, as long as we regard ourselves as agents capable of acting in accordance with a norm – we shall not be able to dispense with the intentional framework of beliefs, wants, and volitions. This again sounds Kantian: our commitment to the intentional framework is a reflection of our nature as rational agents, and our need for it arises out of the demands of practical reason, not those of theoretical reason.