

---

# INDIVIDUATION AND CAUSATION IN PSYCHOLOGY\*

BY

TYLER BURGE

---

Over the last decade I have argued that certain relations between an individual and his environment are partly determinative of what it is for the individual to have certain kinds of mental states and events. In any full explication of the nature of such mental states, such relations will be cited. I call this view “anti-individualism”.

I have grounded this view on a series of thought experiments. I will provide a crude sketch of one of them. One imagines someone *A* with a general familiarity with aluminum but without an ability to provide an account of the nature of aluminum that would distinguish it from every other actual or possible metal. *A* is aware that he is unable to do this, and allows that there might be other metals that would not be aluminum, but which he would be at a loss to distinguish practically or theoretically from aluminum. Nevertheless, *A* does often see, talk about, think about aluminum—as aluminum. He thinks that aluminum is a light metal, for example. *A* is like most of us.

Next one imagines a counterfactual environment in which there is no aluminum and no colleagues of *A* who think or talk about aluminum. This environment contains in aluminum’s place one of those actual or possible metals that *A* could not distinguish from aluminum. Call this metal “twalum”. The environment also contains either *A* or a counterpart of *A* that is for our purposes physiologically identical with *A*, throughout his history. Call this individual “*B*”. (Insofar as there are minor gravitational differences between aluminum and the other metal, I assume that they need not affect *B*’s physiology in any way that is relevant to accounting for his mental states.)

In such a case, *B* clearly does not have any thoughts about aluminum. He does not, for example, think that aluminum is a light metal. He

thinks rather that twalum is a light metal. The difference in thoughts clearly depends on differences in the two individuals' relations to their environments. *A* grew up seeing and otherwise interacting with aluminum. The kinds of thoughts that he can think are determined by these interactions. *B*'s counterpart grew up interacting with twalum, and acquired thoughts grounded in these interactions. Thus the individuation of thought kinds sometimes depends on one's relations to one's environment.

The original thought experiments concerned mental states and events, ordinarily understood. More recently, I showed that non-individualistic modes of individuation can be clearly discerned in the method and theory of cognitive psychology—in particular, the psychology of vision. I think that nearly all parts of psychology that attribute intentional mental states presuppose such modes of individuation for some of their explanatory kinds.<sup>1</sup>

I find this view about psychology natural and unmomentous, once it is realized that intentional states as specified in ordinary mentalistic discourse are individuated non-individualistically. Cognitive psychology makes extensive use of intentional idiom. Nothing critical to the scientific purposes of cognitive psychology seems to be at odds with its sharing with common sense a presupposition of non-individualistic kind individuation. Such individuation supports some of the primary aims of psychology.

Many philosophers do not agree. Although most have conceded the non-individualistic character of mental states as specified in ordinary discourse, many are convinced that psychology cannot reasonably individuate mental states in the same way. Most of the writing from this point of view has not bothered to defend this conviction in any detail. It is devoted to attempts to specify individualistic conceptions of content—"narrow content". These conceptions are not employed outside philosophical circles. Their potential interest derives from the conviction that cognitive psychology (unbeknownst to itself) needs them as surrogates for more ordinary, non-individualistic conceptions. I shall ignore them here, the better to concentrate on the conviction.

I want to discuss a line of objection to my views, developed by Jerry Fodor, that features causation.<sup>2</sup> I hope that this parochial exercise will serve two more general purposes. It will indicate how a common but mistaken view of the special sciences can underlie reasoning about causation. And it will motivate an alternative view of the relation between causation and individuation in psychology.

## I

Fodor's line of objection is very general. It purports to appeal to such general features of science that were it sound, there would be no need to discuss specific theories in psychology. In keeping with the cast of his objection, Fodor has nothing to say to my arguments that the psychology of vision is non-individualistic. The objection begins with a simple argument, and proceeds with some intricate considerations in favor of one of the argument's premises.

The argument is as follows: (1) Psychological explanation is causal explanation. (2) States and processes appealed to in psychological explanation should be type individuated in terms of their causal powers. (3) The anti-individualistic conception postulates the possibility of differences in mental states and processes between two individuals without any corresponding differences in their brain states. (4) But individuals cannot differ in their causal powers if they do not differ in their brain states. So (5) Psychological explanation should not type individuate states and processes anti-individualistically.

I will concentrate on step (4), but first a word about the other steps. Step (3) is acceptable for the sake of argument, but it is not strictly true. The anti-individualistic conception claims that in many cases what it is to be in a certain kind of mental state depends on relations between an individual and his environment. This conception does not entail that two individuals' mental kinds might differ while relevantly corresponding brain states and events remain type-identical.<sup>3</sup> Failure of supervenience of an individual's mental kinds on his neural kinds follows only if relevant differences in the environment do not necessitate differences in the individual's underlying brain states. Now I do reject mind-brain supervenience. I think it "metaphysically possible" for two people with the same brain state kinds (over their whole histories) to have different kinds of mental states. Mental states depend for their natures on relations to the environment in ways that are different from the ways brain states do.<sup>4</sup> I think that the aluminum example can be used to undermine mind-brain supervenience as well as individualism. Thus step (3) is sound, at least *ad hominem*.

Fodor's construal of steps (1) and (2) ignores a widely held position associated with Davidson. According to this position, although psychological explanation is causal in the sense that it makes reference to causal relations, psychological states are not type-individuated to fit causal laws. Psychological explanation provides loose law-like generalizations about events whose causal efficacy is best accounted for only in non-psychological terms.<sup>5</sup>

For the sake of argument I will grant steps (1) and (2) with two qualifications. One is that I will not assume that Fodor and I understand 'causal power' in the same way. The other is that I do not agree that causally relevant type-individuation always focuses on *effects*. The causal *antecedents* of instances of a type (or of events explanatorily associated with instances of a type) are sometimes more significant. Types of rock in geology, species in biology, indicator instruments in science and engineering, optical fibers in physiology, perceptual states in psychology are clearly typed more with a view to causal antecedents than with a view to causal consequents.

The crux of the dispute is step (4) and its notion of "causal power". Before considering Fodor's arguments for step (4), I want to make some background remarks. I assume with Fodor that individuals with the same brain states will make the same movements. I also assume that there are no gaps (into which mental events might swoop) among the events described by the physical sciences. I further assume that if causal chains as described in the physical sciences ("physical causation") did not occur, causal chains as described by psychology ("psychological causation") would not occur.

If one thinks about psychological causation purely from the point of view of "physiological causation" (causation as described by physiology)—or more generally, causation as described in the physical sciences—, one is likely to settle on a use of "causal power" according to which step (4) is unproblematic. Considered apart from individuation in psychology, (4) can certainly seem plausible. But Fodor's notion of causal power is developed in the context of steps (1) and (2). The causal powers of an individual must be understood to be properties that determine how *psychology* individuates its kinds. This restriction is critical in evaluating (4). I shall invoke it repeatedly.

The following point of view motivates accepting (4) in the light of the restriction: Physiological processes are where the "real" causation in psychology goes on. Psychology should concern itself with "real" causation. So psychology should adopt a taxonomic scheme that does not distinguish psychological causes or effects unless there is a distinction in physiological causes or effects. I believe that this point of view drives Fodor's intuitions.

The point of view is certainly odd. It is a precarious strategy to abstract from the aims and practice of psychology in trying to understand psychological kind-individuation. Fodor thinks that he can assume a notion of causal power that is independent of issues regarding psychology and then use it to place restrictions on kind individuation in psychology. I will explain in the following sections why I find this approach hopeless.

## II

I turn now to Fodor's arguments for step (4)—the view that it is impossible for individuals with the same brain states to have mental states with different causal powers. I shall divide the arguments into two stages, letting exposition precede evaluation.

(I) Fodor considers the following view: *A* and *B* have different causal powers because when *A* says the word forms "Bring aluminum", he causes someone to bring him aluminum; whereas when *B* says "Bring aluminum", he causes someone to bring him twalum (since in his environment, those word-forms mean to bring twalum).<sup>6</sup>

Fodor replies that identity of causal powers must be assessed across "contexts", not within "contexts". As an analogy he considers testing the causal powers of his and his reader's biceps: "Roughly, our biceps have the same causal powers if the following is true: For any thing *x* and any context *C*, if you can lift *x* in *C*, then so can I; and if I can lift *x* in *C*, then so can you." If in a context in which a chair is nailed to the floor Fodor cannot lift it, and in another context in which the chair is not nailed to the floor his reader can lift it—that difference does not show that their biceps have different causal powers; for they have been tested in "different contexts". Fodor concedes that when *A* and *B* say the word forms "Bring aluminum", they get different sorts of things in their respective environments. But he claims that this difference is irrelevant to testing their causal powers, because their utterances occur in different contexts. He holds that if the causal powers of *A* and *B* were tested in the same contexts, they would always be the same [pp. 34–35]. I think these claims mistaken. I shall return to them after setting out the second stage of the argument.

(II) Fodor next imagines his opponent's conceding that the causal powers of physiological twins are always the same when tested against effects that are non-intentionally individuated. He considers the view that such twins still differ in their intentionally individuated behavior. One might add that they also differ in numerous other mental effects—including those thoughts, desires, intentions that involve the concept *aluminum*. Fodor offers two replies.

(IIA) He finds it hard to see why the position would not be committed to holding that *A* and *B*'s brain states differ as well as their mental states [pp. 37–38]: *A* is in a brain state that eventuates in his uttering the form of words "Bring aluminum"; so is *B*. If their uttering these forms of words counts as their behaving differently, it looks as though their brain states differ in their behavioral consequences—hence in their causal powers, hence in the brain state types of which they are tokens. But it seems clear that *A* and *B* need not differ in their brain states.

(IIB) The second reply rests on an example that Fodor appeals to repeatedly. He defines "is an H-particle at time  $t$ " so that this phrase applies to a particle at  $t$  if and only if Fodor's dime is heads-up at  $t$ . He defines "is a T-particle at time  $t$ " analogously for tails-up. He points out that whether particles are H-particles has nothing to do with their causal powers. [p. 33]

Fodor thinks that the view that psychology might treat  $A$  and  $B$  as differing in their intentionally specified behavior (and mental states) is analogous to the view that physics might distinguish H-particles from T-particles. But it is hardly obvious that the analogy holds. What Fodor needs to show is that there are no laws or law-like generalizations, of the sort psychology provides, that describe actual causal patterns and that are formulable in the ordinary non-individualistic vocabulary.<sup>7</sup>

Fodor's remarks to this end are not very concise. I will try to summarize them fairly. Fodor states that the trouble with the H-particle method of individuation is that the relations between the coin and the particles are not the right kind to "affect the causal powers" of the latter:<sup>8</sup> "Effects on causal powers require mediation by laws and/or mechanisms; and, in the Twin cases, there are no such mechanisms and no such laws." He supports this claim as follows: For any causal relation between  $A$ 's mental states and instances of aluminum, there must be a corresponding relation that holds between  $A$ 's *neurological* states and instances of aluminum—a sort of causal relation which  $B$ 's neurological states do not enter. Despite this difference,  $A$  and  $B$ 's neurological states are type-identical. This is because "the difference in causal *histories* of their brain states is not of the right sort to effect a difference in the causal *powers* of their brains." He concludes, "Parallelism of argument surely requires us to hold that the differences between the causal histories of the mental states [of  $A$  and  $B$ ] are not of the right sort to effect differences in the causal powers of their minds." [p. 39, note 6]

This supporting argument does not advance the discussion. It simply begs the question. I accept everything in the argument up to the conclusion, but I deny the conclusion. Psychology does not and need not individuate mental states in parallel with brain states.

Fodor says more. He says that on my view about the individuation of psychological kinds, there must be some mechanism that connects the causal powers of an individual's mental states with his environment and that it does so without affecting his physiology. But he claims it is impossible to affect the causal powers of a person's mental states without affecting his physiology:

God made the world such that the mechanisms by which environmental variables affect organic behaviors run via their effects on the organism's nervous system. [pp. 39–40]

You can affect the relational properties of things in all sorts of ways—including by stipulation. But for one thing to affect the causal powers of another, there must be a mediating law or mechanism. It's a mystery what this could be in the Twin . . . cases; not surprisingly, since it's surely plausible that the only mechanisms that can mediate environmental effects on the causal powers of mental states are neurological. The way to avoid making this mystery is to count the mental states—and *mutatis mutandis*, the behaviors—of Twins . . . as having the same causal powers, hence as taxonomically identical. [p. 41]

These remarks risk conflating the question whether causal claims of events that run from the environment to behavior necessarily run "via" neural chains of events, with the question whether *patterns* of causal relation between the environment and the individual could bear on the *individuation* of psychological kinds in a different way from the way they bear on the individuation of neural kinds.<sup>9</sup>

The first quote urges the view that the environment has effects on organic behavior only if it has effects on the organism's nervous system. This view is no objection, since I accept it.

The second passage is hard to interpret. Insofar as it differs from the first, it claims that only a neurological law or pattern can connect environmental effects on an individual with those causal properties of mental states that are relevant to the individuation of mental states. I suspect that if Fodor were not conflating chains of individual events with patterns—chains of types of events—or with laws, he would not have made this claim. For there is no reason why a psychological law or pattern cannot make the connection. Nor is there any reason why a law or pattern involving psychological kinds cannot relate the individual to his environment. Fodor suggests that there is some mystery about failures of supervenience between psychological and physiological kinds, without explaining what it is.

In order to understand better how failures of supervenience of the mental on the physiological are compatible with psychology's assumption that gapless neurological chains must connect stimulus with response, let us set psychology aside for a moment and look at some other sciences. In geology, land masses are typed as plates because of their relations to other land masses. If there were no sliding of land masses across the face of the earth, land masses would not have been typed as plates. Moreover, the causal powers associated with these land masses would have been differently described by geology. But the chemistry and physics of the relevant land masses need not (metaphysically need not) have been affected. A land mass with substantially the same non-relational physical features could—because of its different relations to the environment—be of a different geological kind. Thus geological kinds do not supervene on the kinds of masses that are described by physics and that constitute the geological entities.

Similarly in biology, organs in the body are typed because of their function in the "bodily environment" that surrounds them. Something is a heart because its organic function is to pump blood in a circulatory system that extends beyond the surfaces of the heart. One can imagine an organ in a different sort of body with a totally different function (it might pump waste for example). The causal powers attributed to such an organ by biology would be different from those attributed to a heart. Such an organ would not be a heart, but it might be chemically and structurally homologous to a heart. The biological kind *heart* does not supervene on the chemical structures of material that constitutes hearts.

The special sciences track causal patterns that cut across one another. One science's typology may be more "environmentally" dependent than another's, or dependent on a different sort of environment. In such cases, it is possible for that science (geology, biology) to make its methods of kind individuation sensitive to possible differences in individual-environmental patterns of interaction, while the typology of a science (chemistry, physics) that deals with the underlying physical constitution of the individual remains insensitive to those possible differences.

In my view, despite significant differences between psychology and the natural sciences, psychology's relation to neurophysiology is similar, in the relevant respects, to the just mentioned relations between natural sciences. There are causal patterns between environment and individual that warrant typing certain psychological states in terms of aspects of the environment. Psychology may try to explain such environment-individual interaction. Or it may presuppose such causal relations in its typology, using this typology in explaining the individual's behavior and formation of mental states. (For more on this distinction, see section IV.) Whenever psychology attempts to explain an individual's behavior on the basis of such a typology, there must be neurologically typed causal chains connecting stimulations of bodily surfaces and bodily motions.

But these typologies cut across one another. If the environment had been relevantly different, the patterns of interaction between individual and environment—which are either explained or presupposed—would have been different. The psychological kinds that the individual would instantiate would have been different. But there is no reason in metaphysical principle why these differences must affect the typology of neurophysiology. The patterns of causation that neurophysiology deals with have less to do with individual-environmental relations than those that concern psychology.<sup>10</sup>

This completes my discussion of Fodor's defense of line (IIB). It also constitutes a reply to his line (IIA). Brain sciences are concerned with patterns that are not as sensitive to individual-environmental relations as the patterns that concern psychology.



### III

I turn now to Fodor's first (I) line of defense of the idea that individuals cannot differ in their causal powers if they have the same brain states—step (4) in his argument. It will be recalled that Fodor claimed that causal powers of two individuals must be tested by considering effects of their activities in the same contexts. He further claimed that if the twins were tested in the same contexts, they would always produce the same effects. He took this result to indicate that they have the same causal powers—powers in terms of which psychology individuates its kinds. So the Twins are supposed to be psychologically identical.<sup>11</sup>

The appropriate applications of the notions of causal power and context are, however, less obvious than Fodor supposes. In the first place, there are clearly "contexts" in which the twins' utterances would produce different non-intentionally specified effects in the physical world. Suppose a context in which hearers understand what both *A* and *B* say, wish to comply with their requests, and are in a position to do so. (For no explicit reason, Fodor uses only contexts in which the remarks by exactly one of the twins will be misunderstood by his hearers.) When *A* makes a request with the sounds "Bring aluminum", he is brought aluminum. When *B* makes a request with the sounds "Bring aluminum", he is brought twalum. Since the individuals have different physical effects in the same context, by Fodor's own test they have different causal powers.

The case relies on hearers' understanding. I think that this reliance is appropriate. Hearer understanding is the normal situation when the effects of speech are being considered. Regularities associated with successful communication are presupposed in most experiments in cognitive psychology.

This reliance is, however, inessential. There could be a device that traced the histories of individuals, recording whether they had been in causal contact with aluminum. Such a device could bring aluminum to an individual with such a causal history when he made the sounds "Bring aluminum"—and not otherwise. In such a context, *A* would have different effects from *B*. Once again, there is a possible context in which the twins' acts produce different effects. Unless some restriction is placed on admissible contexts, Fodor's test will count any two individuals with any differences at all in their physical histories as having different causal powers.

Of course, one might justly protest that many such devices will not be keyed to the sort of causal powers that are relevant to causal explanation in science. But this protest brings out a weakness in Fodor's test. The test does not provide an independent check on whether two individuals will

fall under the same explanatory kinds for any given science. Interpreting the test in such a way as to make it individuate those causal powers that are relevant to scientific typology requires that one already know what counts as an admissible kind. But that is precisely the point at issue. Not all historical considerations are irrelevant to typing in the sciences. Whether a rock is igneous, whether an organ is a heart, whether a neural fiber is optical depends on how it was formed and/or what its evolved function is—not merely on how it will respond physically to any conceivable stimulus. Similarly, it is hardly obvious that a context in which a speaker's utterances are understood is an inappropriate context for checking causal powers relevant to psychological explanation. By itself Fodor's test provides no restriction at all on kind-individuation in the sciences.

There is another thing wrong with the test. It is insensitive to differences among the typologies of different sciences. So it is insensitive to the different ways in which sciences recognize causal powers. Consider the case of a heart again. To be a heart, an entity has to have the normal, evolved function of pumping blood in a body's circulatory system. One can conceive of a physically homologous organ whose function is to pump waste—or even a physically homologous entity that came together accidentally and lacks a function. Such entities would not be hearts.

The causal powers of a heart that are relevant to its being typed a heart concern its role in circulating blood. Similarly, many of the physiological laws that govern a heart have to do with its relations to its environment—the relation between heart pressure and pressure in the blood vessels, the relation between regulatory systems in the brain and pumping action by the heart, and so on.

We can imagine that a physically homologous organ whose function is to pump waste is surgically replaced by a heart. This is a physically and perhaps even biologically possible situation. In such a context, the heart would have the same physical effects as its physically homologous counterpart. We can also imagine a heart being surgically replaced by the organ for pumping waste; again that organ would pump blood in the same way that the heart would.

But it would be ludicrous to argue from these cases that the heart and its counterpart have the same causal powers *as typed by physiology* and that there is therefore no difference in kind. From the point of view of some sciences, the two entities would indeed count as type identical. But the physiological differences are patent. Physiology recognizes causal powers of the heart which are exercised in its functionally normal environment. It individuates its kinds with a view to understanding regularities in an actual environment. Of course, other "environments" for the heart, in which it would have very different sorts of effects, are metaphysically possible. The surgical example indicates such possibilities.

Physiology may explain how the heart would act in some of these environments. But these environments are irrelevant to the scheme of kind individuation that physiology actually uses. Fodor's test is insensitive to this dependence of many special sciences on a normal environment for picking out those causal powers that are relevant to an explanatory typology in those sciences.

The analogy to psychology should be obvious. Psychological states are type-individuated to account for causal patterns that are played out in the environment of the individual subject. The fact that from the point of view of physics or physiology, the bodies of *A* and *B* may have the same effects if each were transplanted into the other's context does not show that those causal powers that are of interest to psychology should be counted the same. Insofar as Fodor's test ignores the environmental background against which the individuals' psychological states are type-individuated, it is useless as a test for those causal powers that are relevant to type individuation in psychology.

I conclude that Fodor's line of defense (I) for step (4), like the lines (IIA) and (IIB), fails. I see no other arguments in favor of this step, nor does the step seem plausible in view of the anti-individualistic considerations that count against it.

#### IV

An implicit source of the conviction that there is something scientifically impossible about the individuation of psychological states in an environmentally sensitive way is a misconception about the relation between individuation and environment-individual causal interaction. It is often thought that since bearing some complicated causal relation to aluminum or to water is clearly not a property cited in psychology and is not something that psychology is likely to try to explain, the anti-individualist view must be mistaken.<sup>12</sup>

But a conflation underlies this reasoning. Not every relation between individual and environment that is necessary, sufficient, or importantly contributory to the individuation of mental kinds is itself a candidate for a scientifically useful kind or relation. It is one thing to presuppose a property or relation in one's typology. It is another to use it in one's explanations. Thus a full philosophical understanding of what it is to have the concept of water might refer to causal relations between individuals and their environments. But it is not incumbent on psychology, or any other science, to cite those relations as explanatory kinds. The property *bears such and such causal relation to H<sub>2</sub>O* is not the sort of property that is appropriate for purposes of scientific explanation. The relevant psychological kind is *having the concept water*.

This conflation of scientific kinds with properties that figure in the explication of what is involved in being an instance of those kinds plays a part in an argument originally put forward by Ned Block and repeated by Fodor.<sup>13</sup> Block describes a psychology of individual differences regarding taste. In accordance with one of my examples, we are to imagine twins with different concepts—someone who mistakenly thinks that brisket can only be beef, and a twin who has a concept other than brisket (call it *tw-brisket*) that correctly applies only to beef brisket. Fodor and Block lampoon the view that a psychology of individual differences would attribute a difference between the twins and ascribe the difference to a difference in “linguistic affiliation”, which is the putative source of their conceptual differences. They hold that the science would attribute any differences either to genetic endowment or to early learning.

I see no reason to disagree. The objection does not clearly conflict with anything I hold. Let us grant, what seems dubious, that this branch of psychology is acceptably clear about how it should treat the case. There is no reason to think that the science makes full use of intentional notions. The science appears to be aimed at something like individuals' *de re* preferences among tastes, regardless of the details of how individuals conceptually associate those tastes with food types. Presumably, any given piece of food, regardless of whether it is thought of as brisket or as *tw-brisket*, would taste the same to the twins. I do not doubt that the individual's precise way of thinking about the food taste is a matter of indifference to the science. In such a case, the individuals would be treated as having the same preferences. Some aspects of psychology will be insensitive to some or all the twin-earth thought experiments. Parts of psychology that make do with limited or no intentional notions will fall in this category.

My main reason for discussing the objection derives from another problem with it. Suppose (I do not know how plausibly) a science that concerned itself with distinctively *intentional* differences among food preferences. Thus it would recognize the possibility that an individual might prefer *A* to *B* even though, unbeknownst to him, *A* and *B* were the same food or food taste. Suppose that in such a case, the twins were treated as having different preferences. Contrary to Block's objection, it is not a consequence of my view that such a science need cite linguistic affiliation as the explanatory source of the difference. One would think that it is a consequence only if one conflated scientific explanatory kinds with relations that figure in the account of how those kinds are type individuated. Linguistic affiliation (or the like)<sup>14</sup> need not compete as an explanatory kind with the kinds cited by Block. Innate equipment and early training are the rather crude explanatory categories cited. If the conceptual difference is innate, it would be attributed to innate conceptual endowment; if the difference is learned, it would be attributed to

differences in early training. Differences in idiolect might well be associated with individual conceptual differences deriving from early training; they reflect underlying differences in causal patterns. But they need not be cited as an explanatory kind. Whether a property is cited depends on the explanatory objectives of the science, not merely on its presupposed methods of individuation.

Block's objection may suggest the following reasoning: cognitive psychology does not make reference to social relations and does not attempt to explain linguistic or other social interaction; so it cannot make use of intentional kinds that are individuated by reference to social interaction. As I have indicated, the inference is mistaken. Cognitive psychology can presuppose social interaction in the explication of what it is to have a given propositional attitude without making use of explanatory kinds or relations that refer to social matters, and without attempting to explain social interaction. Dependence on others in learning language may be so inescapable that it is a factor in the individuation of intentional states. Yet the details of causal relations, involving interlocutors, that determine the reference and identity of a person's concepts may not be appropriate for any scientific study, much less the study of psychology. Individual psychology can idealize away from these interactions and make use of unreduced intentional notions as explanatory kinds, while it presupposes the existence of such interactions in the individuation of these kinds.<sup>15</sup>

Even where psychology does attempt to explain individual-environment relations, it need not explain the individuation conditions of its non-individualistic explanatory kinds. This may be seen in the theory of perception, whose concern with such relations is comparatively straightforward. Perceptual theory is largely motivated by explaining the conditions under which perceptions present the physical world veridically (and those under which illusions are created). But perceptual theory does not explain the conditions under which its intentional perceptual kinds are individuated. That enterprise cuts across various scientific disciplines—optics, evolutionary biology, cognitive psychology, and common sense remarks about physical properties and the behavior of observers. Very likely it will remain an enterprise for philosophy and for schematic background remarks within the sciences, rather than for rigorous special-science theorizing.

The general point is hardly peculiar to psychology. The special sciences parcel out their domains of explanation in ways that do not necessarily treat the background conditions that govern their modes of individuation. The physiology of the human heart presupposes but does not explain all the biological conditions that go into making an organ count as a heart. An account of the conditions under which words are individuated may go into historical matters; yet generative linguistics may simply assume

words for a lexicon as already individuated. The social sciences cite psychological kinds, without explaining the individuating conditions for such kinds.

Psychology is partly concerned with how individuals function in the world—how they get around in it; how they act in accord with their conceptions of it; how they perceive it, learn from it, and form certain true beliefs about it; and how they relate to others who function similarly. These individual-environment relations help motivate the explanatory typology of intentional psychology. But psychological explanations of individual-environment relations need not recapitulate, or provide an explication of, the individuation conditions of the kinds cited in those explanations. Scientific explanation and philosophical explication are different enterprises. Philosophers are more likely to forget this than scientists.

#### V

I want to close by reflecting briefly on the notions of causal power and causation in psychology. I have argued that insofar as we associate a conception of causal power with individuation of psychological kinds, we get that conception not from some model drawn from other sciences, but from the explanations that psychology provides. This point generalizes to other special sciences.

Not all the causal powers of an entity, considered in the abstract or from the point of view of physics, are relevant to typing it. The heart has numerous "causal powers" that are irrelevant to its being a heart. It will color a surface red if dropped from a given height; it will damage an eye if it is used to skeet certain solutions at it; it would pump waste if inserted into certain organic systems. None of these powers is relevant to typing something as a heart. None are causal powers recognized by biology or physiology. What are relevant are those causal and receptive powers exercised by the heart that yield the patterns of causation studied in physiology—the powers exhibited when it carries out its basic function, pumping blood in the circulatory system of an organism.

Many of the special sciences, including psychology, study patterns of causation that involve relations between entities and other entities in their normal "environments". Their explanatory kinds are individuated in a way that presupposes such relations. Astronomy studies the motions of the planets; geology studies land masses on the surface of the earth; physiology studies hearts or optic fibers in the environment of a larger organism; psychology studies activity involving intentional states in an environment about which those states carry information; the social sciences study patterns of activity among persons.

Because the kinds recognized by these sciences are individuated by reference to patterns in a normal environment that reaches beyond the surfaces of individuals typed by those kinds, the kinds are not in general supervenient on the constituents of those individuals. It follows that the causal powers *relevant to type individuating the explanatory kinds of a special science* need not be “locally supervenient” on causal powers recognized by sciences that deal with underlying constituents. It also follows that there are imaginable cases in which two entities with the same causal powers recognized by the latter sciences, or by physics, may have different causal powers recognized by a higher-level special science. The heart and the organ that pumps waste, and the twins with different psychological states, provide examples.

It is true that entities studied in the special sciences may act or react in certain ways when they are transported outside their normal environments. A heart is still a heart when it is in free fall, or hooked up in a body in order to pump waste. A person would still act on his beliefs and desires on twin earth. Sometimes the relevant special science will have nothing to say about how the entity will behave in the “abnormal” environment: Physiology has nothing to say about gravitational forces on a heart in free fall. In these cases, one can perhaps infer the entity’s would-be behavior from the laws of other sciences, together with descriptions of relevant properties of the entities. In other such cases, the special science might have something to say. Physiology could predict how much liquid waste the heart could pump, given information about input and assuming that the heart’s anatomical condition were not otherwise altered. Psychology can predict how someone will act on twin earth, given appropriate descriptions of the stimuli. But these fanciful situations do not affect the typology of the special science. The typology is fitted to capture actual patterns, actual law-like regularities, into which the relevant entities enter.

The point that those causal powers relevant to individuating the explanatory kinds of a special science cannot be identified independently of the science’s modes of explanation is fairly obvious on reflection. I am inclined to think that the point applies not only to kind-relevant causal powers, but also to causation. Here, however, the issues are more complicated and less clear-cut. Many philosophers have thought that the best strategy for understanding causation involving psychological events is to assume a notion of causation that is derived from other sciences, and then to shape the interpretation of psychological causation to fit that notion.

The most deeply imaginative execution of this strategy is Donald Davidson’s.<sup>16</sup> Davidson holds that attribution of causal relations entails commitment to a certain sort of explanatory law, a sort of law that has properties (those of being exceptionless and of forming a closed system

of explanation) that one cannot reasonably expect the principles of psychology to exhibit. Mind-body causation is then interpreted in the light of this assumption. Such causation is held to fall under purely physical laws. I think that there is no such entailment. Insofar as causal ascriptions entail commitment to any sort of law, it is an empirical question what sort. Indeed, what counts as a law is filled out partly through scientific practice. One cannot know a priori that every causal relation, regardless of domain, must fall under laws that have any particular form. So Davidson's argument that physical descriptions provide the fundamental insight into causal relations involving psychological events seems to me unpersuasive.

Earlier I cited a cruder version of the same strategy. The argument went: Physiological processes are where the "real" causation in psychology goes on. Psychology should concern itself with "real" causation. So psychology should adopt a taxonomy that distinguishes psychological causes or effects only if in all possible circumstances there are associated differences in physiological causes or effects. In sections I-IV I argued that the inference in the last step is faulty. (Davidson would agree.) But the first premise is doubtful as well. What is true is that physiological processes are *sine qua non* for psychological processes. But atomic and quantum processes are *sine qua non* for physiological processes. Few would advance the view that atomic or quantum processes are where the "real" causation goes on in physiology.<sup>17</sup> Is there any strong ground for thinking that the "true nature" of causation involving psychological events is better revealed in physiology than in psychology?

Our knowledge of mental-physical causation derives primarily from mentalistic explanation. I think that it may be a mistake to seek to reinterpret psychological causation in non-psychological terms. The relations between mental and physical descriptions are more a subject of speculation than of scientific certainty. It is an empirical question what sorts of laws or principles are to be associated with attributions of causal relations. Understanding psychological causation is at least as dependent on what sorts of explanations we achieve in psychology, and how they are related to explanations in the biological sciences, as it is on any antecedent conception of causation. It is therefore an open question whether it will ever be illuminating and correct to count relations between neural events (tokens) as revealing the nature of causal relations involving intentional psychological events. The anti-individualistic position does not depend on rejecting this common view of psychological causation. But I think that the view warrants skepticism.

It is usually a mistake to allow ontological preconceptions that have primarily philosophical underpinning to affect one's interpretation of scientific enterprises. It is almost always a mistake, and a larger one, to allow them to dictate the sorts of explanatory kinds that are deemed



admissible for explanation. I think that many of the objections to anti-individualistic modes of kind individuation in psychology rest on these mistakes. Mentalistic explanations dominate large reaches of psychology. Certainly our notion of psychologically relevant causal power, and probably even our notion of psychological causation, are best illuminated by reflecting on these explanations. Most intentional kinds attributed in these explanations are individuated in non-individualistic ways. These methods of individuation are consonant with the aims of psychology and indeed those of many other special sciences.

University of California, Los Angeles  
Los Angeles, California

#### NOTES

\* I am grateful to Ned Block and David Braun for valuable criticisms.

<sup>1</sup> Cf. my "Individualism and the Mental," *Midwest Studies*, IV, 1979, pp. 73-121; "Other Bodies," in *Thought and Object*, Woodfield, ed., Oxford University Press, Oxford, 1982; "Cartesian Error and the Objectivity of Perception," in *Subject, Thought, and Context*, Pettit and McDowell, eds., Oxford University Press, Oxford, 1986; "Intellectual Norms and Foundations of Mind," *The Journal of Philosophy*, 83, 1986, pp. 697-720; "Individualism and Psychology," *The Philosophical Review* 95, 1986, pp. 3-45.

<sup>2</sup> Jerry A. Fodor, *Psychosemantics*, MIT Press, Cambridge, Mass., 1987, chapter 2. I will cite page numbers from this work in the text.

<sup>3</sup> Here and elsewhere, I understand the notion of "might" or possibility involved in step (3) to be not scientific or physical possibility but something like what is commonly called "metaphysical possibility". (Similarly for "cannot" in step (4).) My views do not depend on exactly how one construes this notion as long as one allows possibilities or illustrative stories that allow conditions of individuation to vary, where these conditions may include physical law. I think that some psychological kinds depend for their natures on physical laws relating individual to environment. In testing this view, one must consider "possibilities" in which physical laws are not held fixed. It is not strictly necessary that one consider such variations of physical law "possible" (although I do incline toward thinking of them as possible) as long as one allows oneself heuristic devices that enable one to bring out the sensitivity of psychological kinds to environmental conditions, including environmental laws. I use the phrase "metaphysical possibility" with these qualifications in mind.

<sup>4</sup> I take it that to *be* a brain involves evolving in a certain way and having certain functions within the environment of the rest of the body, and probably within a larger environment. So an entity that coalesced by quantum accident but which was atomically homologous to a brain would not be a brain. So the natural kind *brain* is itself not supervenient on the atomic make-up of a brain.

<sup>5</sup> Cf. Donald Davidson, "Mental Events," in *Essays on Actions and Events*, Oxford University Press, Oxford, 1980. This view seems to lie behind the excellent discussion of Fodor's argument in Robert Van Gulick, "Metaphysical Arguments for Internalism and Why They Don't Work," in *ReRepresentation*, Silvers, ed. (Synthese Philosophy Series), Kluwer, 1988. I agree with much of Van Gulick's discussion. In particular, we agree that Fodor provides no good reason to believe that individuation of psychological kinds must supervene on individuation of neural kinds. But I want to mention a point of difference. Van Gulick apparently holds that physical descriptions are the primary ones in indicating

causal relations. He apparently does not see psychological principles as causal principles, but calls them psychologically interesting generalizations that, in some unspecified way, throw light on causal patterns. This view may be correct. But I do not myself embrace it because I do not think that there is any clear, well-established sense in which physical descriptions of mental-physical causal relations provide the primary insight into causal relations. Cf. section V.

<sup>6</sup> Fodor discusses other examples I have given. I will key his discussion to the aluminum example. I think that this will not affect the substance of the discussion.

<sup>7</sup> I am using my terminology here rather than Fodor's. Oddly, he characterizes individualism as the view that psychological states are individuated with respect to their causal powers [p. 42]. I agree with this view, with the provisos about the understanding of "causal power" and about the role of causal antecedents in individuation that I have indicated.

<sup>8</sup> Elsewhere, Fodor seems to suggest this definition: "x's having property P affects x's causal powers just in case x wouldn't have caused the same events had it not been P" [p. 38]. He seems to interpret "x wouldn't have caused the same events had it not been P" as "in some possible context, if x had not been P, x would not cause some kinds of events that it now would cause in that context".

<sup>9</sup> Fodor's terms "mechanism" and "affect the causal powers" tend to engender the conflation described above. Sometimes "mechanism" can mean "causal chain of particular events"; sometimes it can mean "law" or "law-governed pattern of causally related events". "Affect the causal powers" can seem to describe a relation between particular events that are the effects of the environment on the individual and particular states or properties that are further effects of these effects. But insofar as it is relevant to individuation, it must be used to describe the bearing of patterns of causal relations (e.g. between the environment and the individual) on the individuation of kinds relevant to causal explanation in psychology.

<sup>10</sup> I made these points in "Individualism and Psychology", *op. cit.* (The quote that I take from this article below is on p. 16.) I warned against confusing causation—a relation among particular events or states—and individuation, which is based on patterns among causal relations. I pointed out that there is a *non sequitur* in the following reasoning: The world is so constituted that the only way that the environment can have effects on an organism's behavior is to have effects on its nervous system, which in turn produce bodily motions that constitute behavior. So any difference in the typology of mental states or behavior must have a corresponding difference in the typology of neural states. This reasoning leaps from necessary dependence of psychological causation on neural causation and continuity of causation at the neural level, to supervenience of psychological kinds on neural kinds. The gap in reasoning is illustrated in the examples from other sciences that I provided.

I had written, "Variations in the environment that do not vary the impacts that causally 'affect' the subject's body may 'affect' the individuation of the . . . intentional processes he or she is undergoing. . . . It does not follow that the environment causally affects the subject in any way that circumvents its having effects on the subject's body". Fodor replies,

But it looks to me like that's precisely what *does* follow, assuming that by "causally affecting" the subject Burge means to include determining the causal powers of the subject's psychological states. You can't both individuate behaviors in Burge's way (viz. *nonlocally*) and hold that the causal powers of mental states are locally supervenient. [p. 41]

This reply is confused. I meant by "causally affects the subject" "*bears causal relations to events or states in the subject's mind or body*". I do individuate some behavior nonlocally, but I take it that the causal powers of mental states—and more generally, the sorts of properties that psychology is concerned with when it gives causal explanations—are

also not locally supervenient; the individuation of the mental states and their causal powers presupposes certain relations between the individual and his environment. Causation is local in the sense that each causal series of particular events between the environment and the individual's mental states necessarily "runs via", or presupposes the existence of, some causal relations (causal series of particular events) between the environment and the surfaces of the individual's body; and each causal relation between the individual's mental states and his behavior is necessarily associated with some causal series between the individual's neural states and his bodily motions (ignoring behavior that does not involve motion). But individuation of mental kinds and behavioral kinds, and individuation of those causal powers associated with these kinds, are nonlocal.

Fodor does get around to questioning this position. But his questioning amounts to unsupported expostulation: "But is Burge seriously prepared to give up the local supervenience of causal powers? *How could* differences of context affect the [individuation of] causal powers of one's mental states without affecting the [individuation of] states of one's brain?" [pp. 41–42] (I have interpolated the phrases in the brackets to guard against further confusion of causation with individuation.) The answer to the first question is "yes". As regards the second, the examples from other sciences show how differences in environmental "context" could affect psychology's individuation of kinds and causal powers without affecting the individuation of kinds or causal powers by a science concerned with the physical constitution of the individual.

<sup>11</sup> The mental events of the twins in the relevant thought experiments cause different mental events and intentional actions. Our discussion has shown how this is possible. But for purposes of criticizing this line of objection, I shall not rely on this view.

<sup>12</sup> Fodor seems to reason this way. He ridicules the anti-individualist conception by repeatedly comparing the property of being an H-particle with that of being a mental state of a person who lives in a world where there is XYZ rather than H<sub>2</sub>O in the puddles [p. 34]. He then points out that living in a world where there is H<sub>2</sub>O in puddles is clearly not a scientific kind. A boring mistake in this reasoning is that, on my view (as on virtually anyone's), just living somewhere has nothing directly to do with having any particular concept, and thus is irrelevant to psychology. (Fodor knows this—cf. p. 39, note 6; but he backslides in the interests, I suspect, of rhetorical point-making—cf. p. 40 and p. 40, note 7.) A more interesting reason for the failure of this sort of objection is the subject of the ensuing paragraphs. The sort of conflation that I shall be discussing also dominates Fodor's "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology," in *Representations*, MIT Press, Cambridge, Mass., 1981, pp. 246–252.

<sup>13</sup> Ned Block, "Advertisement for a Semantics for Psychology," in French et al. eds., *Midwest Studies in Philosophy*, University of Minnesota Press, Minneapolis, 1986. Fodor's discussion of the argument occurs in *op. cit.*, p. 40.

<sup>14</sup> "Linguistic affiliation" like Fodor's "living in a world where there is H<sub>2</sub>O" (cf. n.12) is rather misleading as an indicator of the sort of property I would cite in an explication of what in early training is relevant to conceptual differences. The way that interaction with other people affects one's idiolect is a complex matter. Cf. my "Wherein Is Language Social?" in A. George, ed., *Reflections on Chomsky*, Blackwell, Oxford, 1989. "Linguistic affiliation" suggests primitive ways of thinking about linguistic community.

<sup>15</sup> Several philosophers have held that my anti-individualistic views may be applicable to cognitive psychology, but only insofar as they do not appeal to social interaction. I find this distinction misconceived. It is true that cognitive psychology may attempt to explain (say) perceptual interaction with the environment and may avoid concerning itself with social interaction. But this distinction in explanatory objectives does not entail that there is a parallel distinction in the presuppositions of the explanatory typologies. An individual's thoughts about arthritis can be cited in explanations that idealize away from social interaction, even though an individual's having those thoughts may be explicable partly in terms

of his dependence on others. Cf. "Individualism and the Mental" *op. cit.* and "Wherein Is Language Social?" *op. cit.*

<sup>16</sup> Donald Davidson, "Mental Events", reprinted in his *Essays on Actions and Events*, Clarendon Press, Oxford, 1980.

<sup>17</sup> Ned Block makes substantially the same point in "Can the Mind Change the World?" forthcoming in George Boolos, ed., *Meaning and Method: Essays in Honor of Hilary Putnam*, Cambridge University Press, Cambridge. John Heil and Alfred Mele also make the point in "Mental Causation," unpublished. Incidentally, one should not infer that psychology cannot provide a notion of causation because it does not promise the bump and grind conception suggested by classical mechanics. Such a conception is inapplicable to large reaches of physics.