X*—EXTERNALISM AND MENTAL CAUSATION

by Pierre Jacob

I

Two commonsense theses about mental causation. There are, I think, two commonsense theses about mental causation: a weaker thesis and a stronger thesis. According to the weaker thesis, propositional attitudes are causes of intentional action and they are involved in the causal process leading to the formation of new propositional attitudes. So my intention to raise my left hand is a cause of my intentionally raising my left hand. My intention to drink a glass of orange juice together with my belief that there is orange juice in the fridge can cause me to open the fridge. If upon inspecting it, I discover that the fridge does not, as expected, contain orange juice but contains lemonade instead, I will give up my former belief and I might change my intention to drink a glass of orange juice into an intention to drink a glass of lemonade on account of my newly acquired belief that there is lemonade, not orange juice, in the fridge. If token physicalism is true (as I will assume it is), then the weaker causal thesis is true: if tokens of propositional attitudes are brain state tokens and if brain state tokens can be causes, then so can tokens of propositional attitudes. Token physicalism secures the view (made philosophically respectable by Davidson 1963) that propositional attitudes can be causes or enter causal relations.

According to the stronger thesis, tokens of propositional attitudes are causes of intentional action and they are causally involved in the formation of new propositional attitudes *in virtue of their contents*. On the stronger thesis, not only are mental state tokens causes; in addition, content properties are causally efficacious.

*Meeting of the Aristotelian Society held in the Senior Common Room, Birkbeck College, London, on Monday, 9th March, 1992 at 8.15 p.m.

I will assume that types of propositional attitudes are mental properties. If token physicalism is correct, then mental properties are not expressible (nor definable) by means of predicates standing for basic physical properties. My instantiation of, e.g., the belief that London is pretty is my tokening of a brain state. Such a brain state token will have or instantiate, not only the mental property of being *the belief that London is pretty*, but also physical properties (e.g., electrical and chemical properties). Assuming, as all token physicalists do, that physical properties of an individual's brain are causally efficacious, then the question naturally arises whether mental properties can be, as common sense takes them to be, causally efficacious too. If not, then even though mental state tokens are causes, mental properties, unlike physical properties of an individual's brain, may well turn out to be epiphenomenal or lack causal efficacy.[1]

Although, following Davidson, I will assume that *causation* is an extensional non-epistemic relation between states or events, it still remains an open question which *properties* of a cause are *causally efficacious*. Besides, I will distinguish the causal efficacy of a property from its role in a *causal explanation*. Explanation, unlike causation, is an epistemic notion subject to pragmatic constraints. Even though a property of a cause may turn out to lack causal efficacy, it may still be relevant to causal explanation.

II

Two epiphenomenalist threats. I wish to distinguish two epiphenomenalist threats—a milder and a stronger threat—depending on one's assumptions about how mental properties relate to physical properties of an individual's brain, which in turn depends on one's assumptions about content individuation. Suppose that an individual's mental properties *supervene* on the physical properties of his or her brain. Then the milder threat is that the causal efficacy of mental properties be (in the words of LePore & Loewer 1987, 1989) *screened off* (or

1 See e.g., Horgan (1989), Kim (1984) or Sosa (1984).

pre-empted) by the causal efficacy of physical properties of an individual's brain.

If the mental can be multiply realized by the physical, then two different individuals—say, an English speaker and a French speaker—whose brains have distinct physical properties may safely be ascribed a single mental property with one and the same content—e.g., the belief that London is pretty. The milder threat is that the distinct physical properties of the brain of each speaker, not their common mental property, are causally efficacious in bringing it about that e.g., one utters the English sentence 'London is pretty' and the other utters the synonymous French sentence 'Londres est jolie'.

The stronger threat arises from externalism, the view that, not only are mental properties not expressible (or definable) by means of predicates standing for physical properties, but they do not even supervene on the physical properties of an individual's brain. As the celebrated thought experiments of Putnam (1975a) and Burge (1979) show, on our commonsense ascriptions, the contents of propositional attitudes do not supervene on physical properties of an individual's brain. If externalism is the correct view of the individuation of the contents of many propositional attitudes, as I will assume it is, then content is highly relational: the contents of a pair of beliefs simultaneously entertained by a pair of microphysical duplicates may differ from each other for they may involve relations to items belonging to different physical environments or different linguistic communities. The tension between externalism and mental causation arises from the assumption that, unlike content, causation is *local*. As McGinn (1989: 133) has put it,

> what happens at the causal nexus is local, proximate and intrinsic: the features of the cause that lead to the effect must be right where the causal interaction takes place... The causal powers of a state or property must be intrinsically grounded; they cannot depend upon relations to what lies quite elsewhere.

If externalism is correct, then a pair of beliefs simultaneously entertained by a pair of microphysical duplicates may differ from each other as much as one member of each of the following pairs of things may differ from the other: a genuine Churchill autograph and a fake (Dennett 1983: 44); a genuine Picasso painting and a forgery; a genuine $100 bill and a counterfeit; a photograph

representing Bill and a photograph representing Bob, Bill's identical twin brother (Dretske 1990: 7). Presumably, a perfect forgery—whether a forgery of a Churchill autograph, of a Picasso painting, or of a $100 bill—may reflect photons or pass various chemical tests exactly as the original article would. Nonetheless, although they might be physically indistinguishable, the counterfeit and the genuine article have different monetary values. As noted by Dretske (1990: 7), what gives the genuine article greater monetary value than the counterfeit are 'certain historical and relational facts... that do not supervene on the intrinsic physical properties... of the canvas or paper that has this value, the properties that determine the object's causal power'. The historical difference between the original and its imitation might be physically undetectable. The puzzle externalism creates for mental causation is that differences in content derive from different historical origins which might leave no physical traces.

Geach (1969: 72, 99) has christened *Cambridge changes* 'in Socrates' what happened to Socrates when he became shorter than Theaetetus as a result of Theaetetus' physical growth and what happens to him 'posthumously... every time a fresh schoolboy <comes> to admire him'. Adapting Geach's terminology, we might say that externalism generates the threat that an individual's mental properties turn out to be *Cambridge* properties of the individual.

III

The pre-emption threat. Suppose I suppressed the pain in my right foot by swallowing an aspirin. The pill relieved my pain in virtue of its chemical properties—its being composed of acetylsalicylic acid. I explain the process of pain relief by reference to the chemical properties of the pill. I might also supply an alternative causal explanation of the same fact by telling you that I swallowed an *analgesic* pill. Being analgesic is a *functional* property of the aspirin—a property it may share with pills having different chemical properties. Such a functional property, which supervenes on the chemical nature of the pill, is definable as the second order chemical property of having one of the first order (causally efficacious) chemical properties within a disjunctive class of such. Arguably, in the process of pain relief, being analgesic is not

causally efficacious in the same sort of way the presence of acetylsalicylic acid in the pill is causally efficacious.[2] Given that the (first order) chemical property of the pill is causally efficacious, doesn't its causal efficacy pre-empt (or screen off) the causal efficacy of the functional property?[3]

On my view, the threat of pre-emption can be mitigated by distinguishing two ways properties can be relevant to a *causal explanation* or by distinguishing two kinds of causal explanation. Consider two possible explanations of the same *explanandum*: the glass broke because *Bruno* dropped it vs. the glass broke because *somebody* dropped it.[4] Same causal mechanism, but different explanations. The former *explanans* is a singular proposition; the latter is a general proposition. Suppose the former is true, then the latter is true in virtue of the former. Similarly, I submit, a causal explanation of the suppression of my pain may of course proceed by *referring to* (or naming) the chemical properties of the pill I swallowed. This is what Jackson & Pettit (1989, 1990a) call a *process* explanation. It may also proceed by mentioning the functional property of the pill of being analgesic. The latter explanation, which Jackson & Pettit (1989, 1990a) call a *programme* explanation and I would call a functional explanation, proceeds by *quantifying over* a set of causally efficacious chemical properties different pills might have. Again, the latter explanation is true in virtue of the former.

The reason, I think, a causal explanation which refers to a (chemical) causally efficacious property need not screen off (or pre-empt) a causal explanation which refers to a functional property (and quantifies over a set of causally efficacious chemical properties) is that they do not provide the same causal information about the relevant physical (here chemical) process. The former supplies information about the *actual* path of the suppression of pain; the latter supplies more general information about a class of

2 I concur with Block (1990) and Jackson & Pettit (1989, 1990a).

3 This, I take it, is what Kim has recently labelled the problem of 'explanatory exclusion'.

4 Granted, I don't pick out the class of people who drop glasses functionally (as I pick out the class of analgesic substances). But this is irrelevant to the present distinction.

possible unrealized paths. Endorsing the claim that the former automatically pre-empts the latter amounts to accepting what Jackson & Pettit (1990b) have called the 'fine grain preference' according to which information about the details of a causal process should always be favoured over more general information. Along with them and for reasons first pointed out by Putnam (1975b), I do not think that the fine grain preference is justified.[5]

IV

Cambridge properties and causal explanation. As Dretske (1988b: 80) has recently observed, all the properties of a cause do not equally contribute to its causal efficacy. So, although swallowing an aspirin may relieve the pain in my right foot, none of the following properties of the pill are efficacious in the chemical process whereby it relieves pain: its price; the name of the street in which I bought the bottle containing the pill I swallowed; the colour of the letters on the label of the bottle. Not even the geometrical shape of the pill—its being spherical—contributes to the chemical process of pain relief. As pointed out by Dretske (1988a: 32), had it been cylindrical or cubical, it would have relieved my pain all the same.[6]

The stronger epiphenomenalist threat is not that the causal efficacy of physical properties of an individual's brain screens off the causal efficacy of mental properties. Rather, if externalism is correct, then content is *not* a functional property of an individual's brain. It no more supervenes on physical properties of an individual's brain than the price (or the shape) of an aspirin supervenes on its chemical properties. The threat then is that content turns out to be a *Cambridge* property of an individual's brain. Evidently, it cannot be dealt with like the pre-emption threat.

5 See Putnam's (1975b) famous example of a cubical peg 4.5 cm high which can pass through a square shaped hole 5 cm wide but not through a circular hole 5 cm in diameter. Why? The geometrical explanation is simpler and more general than the physical explanation.

6 Having *some* shape or other is not even a property of the aspirin relevant to relieving my pain since the pill might be crushed into powder, lack a geometrical shape, and still swallowing the powder might contribute efficiently to relieving pain.

V

The challenge of externalism and non-causal explanations. If externalism is correct, then common sense errs in ascribing causal efficacy to content. The question now before us is: Is common sense wrong in expecting content to be relevant to causal explanations? Granted, not all good explanations are causal explanations. A tempting thought is that content is relevant to non-causal explanations. Consider questions of the following form: Why is my son a nephew? Why will my wife become a widow upon my death? Why must I write 'r' in succession twice in order to write 'Larry'? Why is the number 4 greater than the number 2?

As Kim (1973, 1974) has argued, the best explanation for why my son is a nephew is to supply the definition of the concept expressed by the word 'nephew' and show that he satisfies it in virtue of e.g., the fact that I am his father and I have siblings. This is true regardless of the fact that he is my son in virtue of a biological process and that he might have become a nephew *after* I acquired siblings. Similarly, relevant responses to the other three questions will be conceptual explanations: they will exhibit conceptual (not causal) dependencies. Such *explananda* do not call for causal explanations any more than why some theorem of arithmetic is true calls for a causal explanation. Why a given mathematician *believes* the theorem to be true on the other hand is an *explanandum* for a causal explanation.

Some questions may be ambiguous as to whether they are requests for a causal or for a conceptual explanation. Consider the question: Why do I live in Paris? I may point out that I live in Paris because I moved from Lille to Paris (or because I got a job in Paris). Alternatively, if the issue is whether I can vote for some representative of Paris, I may point out that I live in Paris in virtue of living in the third *arrondissement* of Paris. The former is, the latter is not, a causal explanation. What distinguishes causal from non-causal explanations is that the former, unlike the latter, provide information about some physical change or process or that their *explananda* are physical changes or processes.

In order to see the relevance of non-causal explanations to the challenge of externalism, let John and Jack be two microphysical duplicates, whose brains are physically indistinguishable. They can

entertain beliefs with different contents. When John entertains the perceptual belief that cup o_1 in front of him (John) contains coffee, Jack entertains the twin perceptual belief that cup o_2 in front of him (Jack) contains coffee. Let o_1 and o_2 be two perceptually indistinguishable cups containing exactly the same amount of coffee. Now, the truth-conditions of John's and Jack's thoughts differ by virtue of the fact that the former is true if and only if o_1 contains coffee whereas the latter is true if and only if o_2 contains coffee. If, unbeknownst to John, we permute o_1 and o_2 so that John ascribes the property of containing coffee to o_2 which he takes to be o_1, we will be inclined to call his belief *false*, not true. Since they have different truth-conditions, John's and Jack's beliefs are different. Since they differ, they do not supervene on the physical properties of John's and Jack's brains (which by assumption do not differ).[7]

John's belief causes him to drink coffee from o_1. Jack's belief causes him to drink coffee from o_2. Now let us ask the following question: When each seizes the cup in front of him with the fingers of his right hand, brings it up to his lips, tilts his head backwards and drinks coffee from it, does he or does he not do the same thing as his twin? Philosophers sharply disagree. Granted, they perform the same bodily motions. But do they thereby accomplish the same intentional action?[8] If they do not, isn't the *difference* between the contents of their respective thoughts responsible for the difference between their respective actions? If so, content must be relevant to causal explanation.

Perhaps the twins do the same thing, perhaps they do not. Either way, it would be a mistake to assume that whether they do or not is an *explanandum* for a causal explanation. Each twin's belief is a cause of what he does. The twins' actions are two independent twin physical processes. In order to elicit the intuition that two

7 Note that this example clearly shows that externalism is no threat to physicalism: content may well supervene on physical properties of an individual's brain together with physical features of his or her environment.

8 For arguments that they do the same thing, see e.g., Kim (1982), Stich (1983: 160–70). For arguments to the opposite conclusion, see e.g., Evans (1982: 200–204), Hornsby (1986), Peacocke (1981: 198–99).

microphysical duplicates have beliefs with distinct contents, the typical externalist thought experiment sets up a fantastic coincidence between two causally independent processes. The coincidence is not a third physical process above and beyond the two independent processes which constitute it; it is the logical sum of the two separate processes. Bringing out similarities and differences between the respective causes and effects of the two twin processes is not supplying a third causal explanation above and beyond each causal explanation of what each twin separately does. It is a conceptual analysis of the complex coincidence into its conceptual parts. Therefore, although the content of each twin's thought may play a role in the causal explanation of what he does, still the *difference* between the content of John's thought and the content of Jack's thought is not relevant to the causal explanation of the difference between their respective actions.

I therefore take issue with Jackson & Pettit's (1989: 392-93) notion of a programme explanation insofar as it does not discriminate between causal and non-causal explanations. Why impressing two equal forces to two billiard balls of the same mass imparts equal accelerations to each ball is not supplying a causal explanation. Having the same acceleration is not an *explanandum* for causal explanation. That two balls have the same mass, that they are imparted forces of equal magnitude are relational properties of the balls relevant to a conceptual answer to a non-causal question. Jackson & Pettit wrongly, I think, collapse under the notion of a programming property, the explanatory role played by a functional property (like the property of a pill of being analgesic) and the relation of same mass holding between two billiard balls.

I now turn to three kinds of historical relational properties of a system none of which supervenes on the system's physical properties. The question to be addressed is whether they contribute to causally explaining the system's behaviour.

VI

Stich's Replacement Argument. Stich (1978) has argued for what he called the 'autonomy principle' according to which, unless they supervene on physical properties of an individual's brain, properties of an individual will not be relevant to the causal

psychological explanation of his or her behaviour. In defence of this principle, Stich (1983: 165–66) has offered what he called the 'replacement argument'. Suppose I am kidnapped and replaced by an exact physical copy down to the last molecule. Although we are physically and chemically indistinguishable, still there are many things I could do (or could have done) that my physical replica can't do such as divorce my wife or sell my car. What gives me, not my replica, the power to sell my car or divorce my wife is that I bought my car and I married my wife; he did not. Don't we provide a causal explanation of the fact that unlike my replica, I can divorce my wife or sell my car, by mentioning respectively the fact that I, not my replica, stand in the marriage relation to my wife and the fact that I, not my replica, am the owner of my car? Notice that my standing in the marriage relation with my wife no more supervenes on my physical properties—let alone on the physical properties of my brain—than having been painted by Picasso supervenes on physical properties of a canvas. If such legal and historical relational properties of mine can play a role in a causal explanation, why can't content do the same?

Consider what my standing in the marriage relation explains. It does not explain why I divorced my wife, since I did not. Rather, it explains why I *can* (or *could*) divorce her. Even though the fact that I *can* divorce my wife implies or presupposes that I married her—which is a physical process of some sort—still it is not an *explanandum* for a causal explanation since until I divorce her, there is nothing—no process—for anybody to causally explain. However, together with my intention to divorce her, would not my standing in the marriage relation with her contribute to causally explain why I divorced my wife, were I to do so? Similarly, my being the owner of my car explains, not why I sold it, but why I *can* sell it. I may have the right to sell it and never do so in which case there is no *explanandum* for a causal explanation. However, together with my intention to sell it, would not my being its owner contribute to causally explain why I sold it, were I to do so?

Let us look closer at the role played by the property of standing in the marriage relation in the causal explanation of why somebody divorces his wife. Consider the difference between two married men, each having the intention to divorce his respective spouse, one living in a community (e.g. religious) in which divorce is illegal,

the other living in a community in which divorce is legal. The latter, not the former, will be allowed by the relevant authority to divorce his wife if he intends to. The former, therefore, however much he wants to, will not divorce his wife. No divorce, no *explanandum* for a causal explanation. The property of a person of standing in the marriage relation may not therefore enter into a causal explanation unless it is information processed and interpreted by a relevant judge with his or her own beliefs and belief forming capacities. Although relevant to causally explaining a divorce, the information by itself (that the person stands in the marriage relation) is causally inert until it is processed by a human brain.

VII

Dretske's Gizmo. Consider the following example from Dretske (1991). Suppose following an instruction manual you assemble an electro-mechanical device—a 'Gizmo' Dretske calls it—with a given expected output. However, upon turning the electricity on, it does not emit the expected output, it hums and smokes. You want to know why. Given that you know that turning the electricity on is a relevant causal factor in explaining why Gizmo hums and smokes, you want to know more about what within its internal wiring contributes to its humming and smoking. Suppose you discover that a blue wire has been wrongly connected to terminal T_1 rather than terminal T_2. Had you connected the blue wire on to terminal T_2, Gizmo would not hum and smoke. Now, you might have wrongly connected the blue wire to terminal T_1 for at least two different reasons: you might have misapplied the correct instruction manual or you might have correctly applied an incorrect instruction manual.

Now, imagine two microphysically indistinguishable Gizmos *A* and *B* exhibiting the same behaviour when the electricity is turned on: they both smoke and hum. However, *A* smokes and hums because when you assembled it, you incorrectly applied the correct instruction manual, whereas *B* smokes and hums because when I assembled it, I correctly applied an incorrect instruction manual. Obviously, the relation between either *A* or *B* and the instruction manual—a historical relational property of Gizmos—does not, as Dretske (1991: 14) points out, 'supervene on the electro-mechanical constitution of the device whose electro-mechanical behaviour is

being explained'. Isn't that very relation relevant to the causal explanation of *A*'s or *B*'s behaviour?

On my view, it makes sense to split the explanation of the Gizmos' behaviour into two steps. First, we may ask: Why do they hum and smoke? For *A* and *B*, there is one and the same causal answer: Because the blue wire is connected to terminal T_1. The first explanation mentions a physical property of both Gizmos. Second, we may ask: Why is the blue wire connected to terminal T_1? Now, we get a different causal explanation for each Gizmo: one relevant causal factor is a mistake on the part of the person who assembled it; the other relevant causal factor is a mistake on the part of the instruction manual's designer. The reason why I urge that the explanation of either Gizmo's behaviour be split into two explanatory steps is the fact that the electricity being turned on is only relevant to the causal explanation of why the device hums and smokes, not to explaining why the blue wire got connected to the wrong terminal.

Now, in order for either causal explanation of why the blue wire is connected to terminal T_1 to get off the ground, it is not enough to refer to the information contained in the instruction manual (or its relation to either Gizmo). If the instruction manual, correct or incorrect, were just sitting on a shelf, it would do no causal work. We need to refer to your intention to assemble Gizmo, to your ability to understand an instruction manual, to your capacity to form beliefs upon understanding the manual and so forth. If the instruction manual turns out to be incorrect, this fact will be relevant *if* you have correct belief-forming mechanisms and you correctly form what turns out to be the incorrect belief that the blue wire ought to be connected to terminal T_1. Had you, upon incorrectly processing the incorrect information contained in the manual, incorrectly (or perhaps fortuitously) formed the correct belief that the blue wire ought to be connected to terminal T_2, not T_1, you would not have connected the blue wire to terminal T_1.

So unlike the causal explanation of why both *A* and *B* hum and smoke, which does refer to physical properties of Gizmos, the causal explanation of why the blue wire is connected to terminal T_1 does not so refer to physical properties of Gizmos. Nor does it just refer to the abstract information contained in the instruction manual. It also refers both to the intentions and beliefs of the

manual's designer and to the ability to understand the instruction manual, the belief-forming capacity of the person who assembled the device. Obviously, such properties are not properties of Gizmos. They are properties of the brains of at least two separate persons.

VIII

Baker's counterexample to Fodor's necessary condition on causal relevance. In a recent paper, Fodor (1991) has proposed a *necessary* condition for a property of a cause to have relevance in a causal explanation of its effect: if c having property F causes e having property G, for F to be relevant to the causal explanation of why e has G, the connection between c being F and e being G must be contingent, non-conceptual or non-analytic. Consider a pair of microphysically indistinguishable causes and their respective effects: c_1 being F_1 causes e_1 being G_1 whereas c_2 being F_2 causes e_2 being G_2 'only when it is not a conceptual truth that causes which differ in that one has F_1 where the other has F_2 have effects that differ in that one has G_1 where the other has G_2' (Fodor 1991: 19).

Fodor intends his condition to block the difference in *broad* content between the respective thoughts of two microphysical duplicates (e.g., John and Jack) to be relevant to the causal explanation of the difference between their respective actions. On his view, only the common *narrow* content of the twins' thoughts is relevant to the causal explanation of the fact that they do the same thing.

In response to Fodor, Baker (1991) wants to defend the view that (broad) content is relevant to causal explanation and can satisfy Fodor's necessary condition. Consider with her two microphysical duplicates *A* and *B*. *A* lives in an English community in which the word 'jade' denotes either jadeites or nephrites. *B* lives in a counterfactual Burgian community in which 'jade' denotes only jadeites. Suppose that *A* and *B* both appear as contestants in their respective communities, on qualitatively identical quiz shows. For the grand prize each has to identify a stone. The stones, it turns out, are identical pieces of nephrite. Each quiz show host says: 'Here is a lovely green stone. Can you identify it?' To this, *A* and *B* give acoustically identical replies: 'This stone is jade'. At this point the stories depart. For *A* has given the right answer and *B* has given a wrong answer.

In Fodor's framework, c_1 and c_2 are two tokens of the same neural or brain state type, the former instantiated by A in A's community, the latter instantiated by B in B's community. c_1 has mental property M_1 of being a belief with the same truth condition as an utterance of the English sentence 'This is either jadeite or nephrite'. c_2 has mental property M_2 of being a belief with the same truth condition as an utterance of the English sentence 'This is jadeite'. c_1 causes A's utterance u_1 of sentence type 'This is jade' which is true if and only if the stone referred to by 'this' is either jadeite or nephrite. c_2 causes B's utterance u_2 of the same sentence type 'This is jade' which is true if and only if the stone referred to by 'this' is jadeite. u_1 has the property of being a winning answer whereas u_2 has the property of being a losing answer. Baker concludes that content has relevance to causal explanation since the difference in truth condition between causes c_1 and c_2 is relevant to the difference between effect u_1's being a winning answer and u_2's being a losing answer.

On Fodor's necessary condition, for the content of c_1 to be relevant to the causal explanation of why u_1 is a winning answer, the connection between c_1's truth condition and u_1's being a winning answer must not be a conceptual or a necessary truth. In order to demonstrate that the connections are neither conceptual nor necessary, Baker imagines a world in which the same linguistic conditions hold but there is no quiz show. She claims that in such a world, u_1 and u_2 lose their property of being respectively a winning and a losing answer.

Baker's argument faces, I think, the following dilemma: either she treats being a winning answer on the model of truth or she does not. Let us concentrate on the relation between c_1 and u_1.

First horn of the dilemma: she treats being a winning answer on the model of the semantic property of utterances to be correct or true answers. Arguably, an utterance is true (or false) whether or not any judge hears the utterance, interprets it and forms a belief about whether it is true or false. On this horn of the dilemma, being a winning answer is a property u_1 has in virtue of being true. Furthermore, it is a property of u_1 whether or not a quiz show is taking place. One might be tempted to think that there is a conceptual connection between the property of c_1 of having the same truth condition as an utterance of the English sentence 'This

is either jadeite or nephrite' and the property of u_1 of being a winning answer. In which case Fodor's necessary condition would be violated. However, this temptation should, I think, be resisted. The truth *condition* of the thought does not *entail* the truth *value* of the utterance in any standard sense of 'entailment'. Rather, what happens is that being a winning answer is, on this horn of the dilemma, an abstract property conferred upon an utterance by the rules of a game. Uttering a winning (viz., a true) answer, on this interpretation, is not what common sense means by winning the grand prize, i.e., bringing the money home. The rules remain causally inert until the game is played in the presence of a judge. On this view, uttering a winning answer does not secure bringing the money home. Similarly *mutatis mutandis* for the property of c_2 of having the same truth condition as an utterance of the English sentence 'This is jadeite' and the property of u_2 of being a losing answer.

Second horn of the dilemma: assume that, unlike truth and falsehood, being a winning (or a losing) answer is a property an utterance has only if a judge takes it to be correct or incorrect, true or false. This is the sense of 'winning' intended by common sense. Now the causal explanation of why producing u_1 allows A to bring the money home will involve not just properties of c_1 but also the beliefs of the quiz show host and therefore his or her belief forming mechanisms. So on this view, the host must have an appropriate belief forming capacity, which, I take it, is a property of the quiz show host's brain. But then in the quiz show situation, c_1 being M_1 will be neither sufficient nor even necessary for bringing the money home.

First, it is not enough that u_1 be correct or true to be a winning answer. The quiz show host must in addition correctly form the correct belief that u_1 is a winning answer. If he or she incorrectly believes that u_1 is incorrect, then the utterance will not be a winning answer. Second, it is not even necessary that u_1 be a correct answer to be a winning answer. If it were correct, but the judge incorrectly took it to be incorrect, it would not be a winning answer. *Mutatis mutandis* for the relation between the semantic property of c_2 and the property of u_2 to be a losing answer.

In this paper, I have argued that common sense is wrong to expect that content properties are causally efficacious but right to assume

that they are relevant to causal explanations. I have examined the explanatory role played by three kinds of historical relational properties of a system in explanations of the system's behaviour: the marriage relation holding between a man and his wife, the relation holding between an artefact and its instruction manual, and a thought's truth condition. Like content and unlike the functional property of aspirin of being analgesic which supervenes on its chemical properties, such relational properties do not supervene on the physical properties of the system whose behaviour they help to causally explain. I have entertained and rejected the possibility that they are relevant to non-causal explanations. I have argued instead that such properties are relevant to the causal explanation of a system's behaviour via the cognitive activities of some information processing device external to the system whose behaviour is being explained.[9]

CREA
Ecole Polytechnique/CNRS
1, rue Descartes
75005 Paris

REFERENCES

Baker, L.R. (1991) 'Belief in Explanation', mimeo, Middlebury College.
Block, N. (1990) 'Can the Mind Change the World?', in G. Boolos (ed.) *Mind and Method, Essays in Honour of Hilary Putnam* (London: Cambridge University Press).
Burge, T. (1979) 'Individualism and the Mental', in P.A French, T.E. Uehling, Jr. & H.K. Wettstein (eds.) *Midwest Studies in Philosophy*, vol. IV (Minneapolis: University of Minnesota Press).
Davidson, D. (1963) 'Actions, Reasons and Causes', reprinted in D. Davidson, *Essays on Actions and Events*, (Oxford: Clarendon Press, 1980).
Dennett, D. (1983) 'Three Kinds of Intentional Psychology', reprinted in D. Dennett, *The Intentional Stance* (Cambridge, Mass.: Bradford/MIT Press, 1987).
Dretske, F. (1988a) 'The Explanatory Role of Content', in R.H. Grimm & D.D. Merrill (eds.) *Contents of Thought*, (Tucson: Arizona Press).
Dretske, F. (1988b) *Explaining Behavior*, (Cambridge, Mass.: Bradford/MIT Press).

9 I am grateful to Daniel Andler, Lynne Rudder Baker, Frank Döring, Fred Dretske, Paul Horwich, François Recanati and Dan Sperber for criticisms.

Dretske, F. (1990) 'Does Meaning Matter', in E. Villanueva (ed.) *Information, Semantics & Epistemology* (Oxford: Blackwell).

Dretske, F. (1991) 'What Isn't Wrong With Folk Psychology?' to appear in *Metaphilosophy*.

Evans, G. (1982) *The Varieties of Reference* (Oxford: Clarendon Press).

Fodor, J.A. (1991) 'A Modal Argument for Narrow Content', *The Journal of Philosophy*, vol. lxxxviii, No. 1, 5–26.

Geach, P. (1969) *God and the Soul* (London: Routledge & Kegan Paul).

Horgan, T. (1989) 'Mental Quausation', in J. Tomberlin (ed.) *Philosophical Perspectives*, vol. 3, 'Philosophy of Mind and Action' (Atascadero: Ridgeview).

Hornsby, J. (1986) 'Physicalist Thinking and Behaviour', in P. Pettit & J. McDowell (eds.) *Subject, Thought and Context*, (Oxford: Clarendon Press).

Jackson, F. & Pettit P. (1989) 'Functionalism and Broad Content', *Mind*, vol. xcvii, No. 387, 381–400.

Jackson, F. & Pettit P. (1990a) 'Programme Explanation: a General Perspective', *Analysis*, vol. 50, No. 2, 107–17.

Jackson, F. & Pettit P. (1990b) 'In Defense of Explanatory Ecumenism', mimeo, Canberra, The Australian National University.

Kim, J. (1973) 'Causes and Counterfactuals', reprinted in E. Sosa (ed.) *Causation and Conditionals* (Oxford: Oxford University Press, 1975).

Kim, J. (1974) 'Non-causal Relations', *Noûs*, vol. 8, 41–52.

Kim, J. (1982) 'Psychophysical Supervenience', *Philosophical Studies*, vol. 41, No. 1, 51–70.

Kim, J. (1984) 'Epiphenomenal and Supervenient Causation', in P.A. French, T.E. Uehling & H.K. Wettstein (eds.) *Midwest Studies in Philosophy*, vol. 9 (Minneapolis: University of Minnesota Press).

Kim, J. (1989) 'Mechanism, Purpose, and Explanatory Exclusion', in J. Tomberlin (ed.) *Philosophical Perspectives*, vol. 3, 'Philosophy of Mind and Action' (Atascadero: Ridgeview).

Kim, J. (1990) 'Explanatory Exclusion and the Problem of Mental Causation', in E. Villanueva (ed.) *Information, Semantics & Epistemology* (Oxford: Blackwell).

LePore, E. & Loewer B. (1987) 'Mind Matters', *The Journal of Philosophy*, vol. lxxxiv, No. 11, 630–42.

LePore, E. & Loewer B. (1989) 'More on Making Mind Matter', *Philosophical Topics*, vol. xvii, No. 1, 175–91.

Peacocke, C. (1981) 'Demonstrative Thought and Psychological Explanation', *Synthese*, vol. 49, No. 2, 187–217.

Putnam, H. (1975a) 'The Meaning of "Meaning"', reprinted in H. Putnam, *Philosophical Papers*, vol. 2 (Cambridge: Cambridge University Press).

Putnam, H. (1975b) 'Philosophy and Our Mental Life,' reprinted in H. Putnam, *Philosophical Papers*, vol. 2, *op. cit.*

Sosa, E. (1984) 'Mind-Body Interaction and Supervenient Causation', in *Midwest Studies in Philosophy*, vol. 9, *op. cit.*

Stich, S. (1978) 'Autonomous Psychology and the Belief-Desire Thesis', *The Monist*, vol. 61, No. 4, 573–91.

Stich, S. (1983) *From Folk Psychology to Cognitive Science* (Cambridge, Mass.: MIT/Bradford Press).