

MENTAL CONTENTS IN A WORLD OF CAUSES

JOSEP E. CORBÍ
and JOSEP L. PRADES

In everyday life we explain and predict people's behaviour by appealing to their beliefs and desires. Several philosophers, in cognitive science, argue that, if such explanations and predictions are at all possible, it is because we assume a number of generalizations concerning people's mental states and behaviour. Insofar as these generalizations are minimally articulated, we can conclude that our ordinary accounts of people's behaviour presuppose something like a psychological theory, that is, a commonsense psychology. A central feature of this commonsense psychology is that it posits mental states that have causal powers in virtue of their contents. This trait is certainly included in Frege's conception of thinking, since, in his own terms, thinking is grasping thoughts (i.e., a certain kind of content) and thoughts causally affect the world by being grasped by the mind.¹

In any event, one of the fundamental issues in cognitive science is precisely to determine whether mental contents can be causally efficacious and, therefore, preserved in a scientific psychology. In this paper, we intend to pursue some aspects of this fundamental issue by examining Fred Dretske's most recent contributions. Dretske defends the causal efficacy of mental content *qua* mental content², but imposes an important restriction: mental contents can operate as structuring, but not as triggering causes. We shall argue, by contrast, that mental contents can also intervene as triggering causes.

We shall divide the paper into four sections. Firstly, we shall introduce the terms of the discussion, which include two basic principles: the strong supervenience constraint and the intrinsicness condition. Both principles impose restrictions upon the sorts of properties that can be causally efficacious. The intrinsicness condition stresses that only the intrinsic properties of a system can be causally efficacious, while the strong supervenience constraint makes a traditional point about the causal efficacy of properties. In the second section, we shall introduce Dretske's distinction between structuring and triggering causes, and accept that it actually undercuts the intrinsicness condition as a necessary requirement for causal efficacy. Since structuring causes are typically extrinsic. Yet, Dretske assumes that triggering causes still need to respect the intrinsicness condition. The third section is precisely devoted to questioning this assumption. Specifically, we shall call into doubt the strong supervenience constraint that the intrinsicness condition presupposes. Thus, we shall conclude that, despite their essential extrinsicness, mental contents (*qua* mental contents) can operate as triggering causes. We end the paper with a fourth section which recapitulates our fundamental claims.

I THE TERMS OF THE DISCUSSION

Dretske's main goal, in *Explaining Behaviour*, is to show how mental contents can play a role "in the causal explanation of human behaviour."³ According to Dretske, mental contents are explanatorily relevant insofar as they are causally efficacious. But Dretske, as opposed to Donald Davidson, is not interested in the mere causal efficacy of events that possess mental content, but in whether an event can be causally efficacious *in virtue* of possessing a certain mental content. Only if we succeed in accounting for this last possibility, will we be able to recognize the causal efficacy of mental contents *qua* mental contents:

"We can, following Davidson [Davidson, 1963], say that reasons *are* causes, but the problem is to understand how their being reasons contributes to, or helps explain, their effects on motor output.

.....

Trying to exhibit the causal efficacy of meaning one is exploring the possibility, not of meaning itself being a cause, but a *thing's having meaning* being a cause or of the *fact that something has meaning* being a causally relevant fact about the thing."⁴

To appreciate Dretske's undertaking we need first to state why Dretske and other philosophers may find it problematic that reasons should be relevant to the causal explanation of human behaviour. To put it in a nutshell, the philosophical issue⁵ that Dretske addresses is whether the causal efficacy of mental content is compatible with materialism. Let us, then, begin by elucidating Dretske's materialist stance, as opposed to other materialist positions.

Dretske confesses to be a materialist,⁶ but dedicates little time to describing what sort of materialism he is committed to. One can obtain, though, a more precise picture of the main claims that compose Dretske's materialism out of his philosophical practice, that is, by examining the way he approaches particular philosophical problems. Specifically, it seems clear that Dretske's account of the explanatory relevance of mental content is pertinent only if he assumes a strong supervenience thesis of mental properties upon physical properties. We therefore propose to assume that Dretske is committed to the following materialist thesis: All properties in the world are either physical properties or properties that meet the strong supervenience constraint. Following Jaegwon Kim⁷, we can say that the *strong supervenience constraint*⁸ is met by a property when there are certain physical properties that constitute a sufficient condition for the instantiation, in all possible worlds⁹, of this particular property. *Strong materialism* is the sort of materialism that includes the strong supervenience constraint. Correspondingly, mental properties will only be compatible with strong materialism insofar as they meet the strong supervenience constraint. Hence, meeting the strong supervenience constraint is a necessary condition to acknowledge the causal explanatory relevance of mental contents. For only real properties can actually be causally efficacious. Objections have been raised, however, that there are principled reasons against mental contents meeting such a constraint. We shall examine Dretske's strategy to rebut a number of such objections, but, before getting into details, a few more principles and distinctions need to be introduced.

Strong supervenience is opposed to global supervenience. The latter

imposing a very general constraint according to which two worlds which share all their physical properties must also share all their non-physical¹⁰ properties. We should emphasize that global supervenience is stating a constraint that all nonphysical properties need to meet, but is not concerned with all the constraints that any particular kind of nonphysical property must meet. In other words, it is entirely compatible with global supervenience the fact that certain nonphysical properties should be not only globally, but strongly, supervenient upon the physical properties of the world. Accordingly, we define *global materialism* as the sort of materialism that is committed to the global supervenience constraint, but not to the strong supervenience constraint. It seems obvious that the causal efficacy of mental contents is compatible with global materialism. For only a dualist would doubt that two physically identical worlds will also share all their mental properties. Consequently, our philosophical issue only arises if we are concerned with strong materialism.

It has also been argued, however, that, apart from the strong supervenience constraint, a further requirement needs to be met before recognizing the causal relevance of nonphysical properties, namely, the *intrinsicness condition*. This condition demands that only the intrinsic (i.e., individualistically individuated) properties of a system can be explanatorily relevant. The intrinsicness condition presupposes the strong supervenience constraint, since a property is to be judged intrinsic only if it strongly supervenes upon the physical properties of the organism that possesses them. Correspondingly, we shall speak in this case of *intrinsic materialism*.

One may find an obvious motivation for the intrinsicness condition is the intuition that two intrinsically indistinguishable systems will necessarily produce the same outputs, have the same causal powers, no matter how different their relational properties could be¹¹. Two intrinsically indistinguishable cars will perform equally, even if one is ours and the other is yours. If we apply the intrinsicness condition to mental contents, then we are bound to conclude that mental contents can only be explanatorily relevant if they are intrinsic properties of an organism; in other words, if they are in the head. The problem is that mental content may have trouble meeting the intrinsic supervenience constraint, since they appear to be extrinsic or relational properties of

an organism.¹² Thus, Putnam and others¹³ have shown, for instance, that if mental contents are to meet such a constraint they cannot be construed in the traditional guise, that is, as Fregean thoughts. For Fregean thoughts are supposed to fix not only the sense but the reference of our propositional attitudes.¹⁴ But it can be argued that the reference of an organism's mental state is a relational property of it, whereby its determination partly depends on circumstances that are external to the organism. Consequently, two organisms could share all their intrinsic properties and still differ in the reference of their mental contents. This is, though, what the intrinsicness condition excludes. Since everybody tends to accept the validity of this argument, let us assume that Fregean thoughts are not compatible with intrinsic materialism.

People who are committed both to intrinsic materialism and the causal efficacy of mental content certainly need to provide an alternative notion of mental content that squares with the intrinsicness condition. This is the direction the notion of narrow content points to. For years, the development of a plausible notion of narrow content has appeared as the only available strategy to acknowledge the causal relevance of mental content. The results, though, have not been very promising. Some defences of narrow content look rather like desperate moves.¹⁵ Dretske, by contrast, shifts to a different perspective. Instead of seeking to show the compatibility of the causal efficacy of mental contents with intrinsic materialism, he calls into question the intrinsicness condition for the causal relevance of properties.

Accordingly, Dretske's vindication of the causal relevance of mental contents can be divided into two steps: (i) Showing that the intrinsicness condition is not a necessary requirement for the causal relevance of properties and, therefore, challenging the idea that mental contents need to be in the head in order to be explanatorily relevant; and (ii) developing an account of how mental contents can meet the strong supervenience constraint, which still holds.

In the next section, we shall describe the first step in Dretske's approach by introducing his distinction between structuring and triggering causes. This distinction undermines the intrinsicness condition as a necessary requirement for causal efficacy. Since structuring causes can certainly be extrinsic. Yet, according to Dretske, triggering causes

still need to respect the intrinsicness condition. The third section will be devoted to challenging precisely this last conviction. As a result, Dretske's naturalization program (the second step) will become metaphysically superfluous.

II STRUCTURING CAUSES NEEDN'T RESPECT THE INTRINSICNESS CONDITION

Dretske is ready to concede the extrinsicness of meaning. In the debate about narrow content, Dretske undoubtedly takes an externalist position. Now the challenge is to prove that extrinsicness does not deprive mental content of explanatory relevance. As Dretske indicates, the point goes as follows:

"The point, then, is not that extrinsic properties make no causal difference (they can), but that, in order to acquire an explanatory relevance they must supervene on, and work through, intrinsic differences in the system whose behaviour they serve to explain. In order to capture this important feature, therefore, I stipulate that by an extrinsic system I mean a system whose behaviour (some of it anyway) is causally explained by extrinsic (*qua* extrinsic) properties *that do not supervene on the intrinsic properties of the stuff of which the system is composed*. Two such systems can be intrinsically the same, and in this sense physically indistinguishable, and still differ in the extrinsic properties that explain their respective behaviours."¹⁶

The notion of behaviour as an endogenously generated process and the distinction between structuring and triggering causes is at the core of Dretske's attempt to defend the causal efficacy of extrinsic (*qua* extrinsic) properties. According to Dretske, "Behaviour—whether of an animal, a plant, or a machine—is a process in which some kind of external movement [M] or change is produced (caused) by internal events [C]."¹⁷ Schematically, behaviour is a process $C \rightarrow M$, where C is an internal cause and M a movement.¹⁸ Dretske's notion of process does not reduce, however, to a mere regular sequence of events, but is concerned with the bringing about of one event by the other.¹⁹ Behaviour is, then, the process by which C is bringing about M. Accordingly, we must not confuse behaviour with motor output. For behaviour is the

entire process $C \rightarrow M$, while the output reduces to M .²⁰ In any event, if behaviour is a process, how can behaviour be causally explained?

Dretske distinguishes between the triggering and structuring causes of a process. To adapt an example proposed by Dretske himself²¹, suppose I press the arrows on my computer keyboard and the cursor moves in the appropriate direction. I may ask myself *why* this is so. Two sorts of explanatory demands hide behind this single 'why.' Suppose I have been pressing keys at random and do not know the pressure of which one caused the cursor to shift to the right. In such a case, I am interested in getting to know the particular key that provoked such a shift. I am interested in the *triggering cause* of the move of the cursor to the right. In general, the triggering cause T of a movement M is the event of a certain type that, together with certain background conditions, causes M .²² We shall refer to explanations that indicate the triggering cause of a process as *triggering explanations*.

Suppose now that I already know which key on my computer keyboard shifts the cursor to the right. In other words, I know that there is a process which goes from pressing a certain key (C) to a shift of the cursor to the right (M). In seeking to explain this process, I may focus on the triggering cause of C , that is, on the event of a certain type (for instance, my fingertip movement) which caused C , which in turn caused M . But I may want to know why C has got wired to M , i.e., why my computer has got structured in such a way that C causes M . In such a case, Dretske would say that we were looking for the *structuring cause* of the process $C \rightarrow M$.²³ The structuring cause accounts not for C being caused, but for C causing M . The relevance of explanations that appeal to structuring causes is particularly manifest in cases of malfunctioning. Think when the definition of the keys on your keyboard have been changed in order, for instance, to write in another language with a different alphabet. At this juncture, one would certainly be eager to know how C got tied to M and how it could be linked to a different output, so that one's own keyboard should recover its reassuring habitual behaviour. Finally, it seems obvious that the distinction between triggering and structuring causes not only applies to machine processes, but to all sorts of processes, animal and human behaviour included.

If we acknowledge the legitimacy of *structuring explanations* (that is, causal explanations that are concerned with the structuring causes of

processes), we are bound to challenge the intrinsicness condition for the explanatory relevance of a property. For structuring causes are concerned with extrinsic properties of the process they are meant to explain. Two intrinsically identical processes may have quite different structuring causes. For instance, suppose two computers A and B where the pressing of the key labelled 'ñ' produces the sign '*:*' on the screen. The structuring cause in computer A may be that somebody has substituted the American definition of the keyboard for the Spanish one; whereas in computer B the structuring cause of such a process might be the presence of a certain virus.

Objections could be raised that in our example computers A and B were not intrinsically indistinguishable, but only the process $\bar{n} \rightarrow \cdot$. But suppose computers A and B are intrinsically identical and the process ' \bar{n} '—'*:*' still holds in both. We may still have different structuring causes, that is, distinct responses to the question about why ' \bar{n} ' got tied to '*:*'. In computer A the structuring cause may be a mistake in the process of production, whereas in computer B the structuring cause may not be a mistake in the process of production, but in the design itself.²⁴ It seems, then, that two intrinsically indistinguishable processes may have different structuring causes. Consequently, we should abandon intrinsicness as a necessary condition for causal efficacy.

Dretske's argument has a limited scope, though. It certainly demonstrates that extrinsic (*qua* extrinsic) properties can legitimately participate in causal explanations, but only in structuring explanations. In other words, extrinsic (*qua* extrinsic) properties can be structuring causes but not triggering causes. Extrinsic properties could only count as triggering causes insofar as, relative to a given context, they locally supervene upon physical intrinsic properties.²⁵ But, in this case, extrinsic properties would not be, as Dretske points out, explanatorily relevant, since they would not be causally efficacious *qua* extrinsic properties. We can, then, conclude that the intrinsicness condition does not apply to structuring explanations, but only to triggering causes. Let's refer to this limited condition as *the intrinsicness* condition*. We can, then, say that Dretske is committed to *intrinsic* materialism*, that is, a version of materialism which assumes strong materialism together with the intrinsicness* condition.

This conclusion is directly relevant to the causal efficacy of men-

tal contents. For, as Dretske stresses, mental contents are extrinsic properties of an organism whereby mental contents (*qua* mental contents) can only be structuring causes, but not triggering causes, of behaviour. Dretske feels entirely satisfied with this conclusion, since he claims that it captures our essential intuitions about the explanatory relevance of mental content.²⁶ We²⁷ are not completely happy with this restriction and consider that commonsense intuitions point to viewing mental contents as triggering causes as well. Roughly speaking, it seems intuitive to claim that the content of my beliefs and desires frequently act as triggering causes of my outputs. Typically, I reach the fridge because I think there is a beer and I want to have a beer. It seems that, in such cases, my belief and desire, having the content they have, cause my bodily movements. The problem is that these intuitions do not square with Dretske's intrinsic* materialism. What we purport to do next is precisely to vindicate mental contents (*qua* mental contents) as triggering causes by developing an argument which will challenge not only Dretske's intrinsic* materialism, but ultimately strong materialism too.

III EXTRINSIC (*QUA* EXTRINSIC) PROPERTIES AS TRIGGERING CAUSES

III (i) The Principle of Property Causation

We have been speaking so far about the causal efficacy of intrinsic and extrinsic, physical and mental, properties without working out what is the sort of thing that causal relations are supposed to connect. We have seen, though, that the current issue about the explanatory relevance of content is not merely the causal efficacy of mental content, but the causal efficacy of mental content *qua* mental content. It seems obvious, however, that this question can only be raised if we assume that it makes sense to speak of causal relations between events *in virtue* of having certain properties. For what we are discussing is whether an event C, with a content property R, brings about an event E with a property B in virtue of having content property R. Within a Davidsonian approach, where causal relations only hold between events, we

can certainly say that mental events are causes but not in virtue of being mental. Hence, the issue about the causal efficacy of mental content *qua* mental content necessarily assumes what we might call *the principle of property causation*, that is, that there are causal relations between events *in virtue* of having certain properties. In this context, we wish to oppose facts to events, a *fact* being an event E having a certain property P. This is, though, a pure terminological option. All that we say about facts can trivially be described as concerning events in virtue of having certain properties. Hence, those readers that would prefer to say that causal connections hold between events should adequately decode our references to facts. Accordingly, we can now say that what the principle of property causation contends is that there are causal relations among facts.

Intrinsic and extrinsic facts can be distinguished insofar as they are respectively concerned with intrinsic or extrinsic properties of events²⁸. The principle of property causation by itself imposes no constraint upon the sorts of facts (i.e., extrinsic or intrinsic) that can be causally related. Yet, the intrinsicness* condition and the strong supervenient constraint are meant to complement this principle by imposing some constraints. It seems obvious, however, that the intrinsicness* condition can be supported by the strong supervenience constraint. It follows from the strong supervenient constraint that the triggering properties of a system must strongly supervene on the physical properties of the system relative to a given physical context. Consequently, extrinsic properties could only count as triggering causes insofar as, relative to a certain physical context, they locally supervene upon the physical properties of the system. But, in such a case, extrinsic properties would not count as triggering causes *qua* extrinsic properties. Hence, strictly speaking, only the intrinsic properties of a system can count as triggering causes, and this is what the intrinsicness* condition urges.

We are, however, reluctant to concede the strong supervenience constraint. Thus, in the section III (ii), we shall produce an argument to call into doubt this last constraint and, in section III (iii), derive the consequences of our challenge for the intrinsicness* condition and Dretske's naturalization program.

III (ii) The Strong Supervenience Constraint Metaphysically Unmotivated

Not all facts can enter causal explanations. According to Dretske, the liberality of the principle of property causation is conditioned by the strong supervenience constraint. All sorts of explanans and explananda are permitted insofar as they are strongly supervenient upon physical explanans and explananda. Thus, even structuring causes provide causal explanantions only if they strongly supervene upon physical facts and their corresponding triggering connections. To give an example, evolutionary explanations are structuring explanations. We can imagine different evolutionary explanations of the same type of fact, but each of these evolutionary explanations will only count as a causal explanation if there is an implementing mechanism, a set of triggering causes, that constitutes the physical supervenient base upon which evolutionary properties strongly supervene.

Our question is, however: Why should we accept the strong supervenience constraint upon the principle of property causation? We cannot examine in detail all the arguments that have been provided in the recent literature²⁹ to answer this question. We shall concentrate, though, on what may constitute the fundamental intuitions behind the endorsement of the strong supervenience constraint. The argument for this strong constraint goes as follows.

The principle of property causation without the limit imposed by the strong supervenience constraint infringes the principle of physical closure of the world and leads to dualism. For, if we gave up the strong supervenience constraint, we would be bound to concede the existence of nonphysical phenomena that have no physical explanation, that is, that have not been physically caused. This consequence clearly breaks the principle of physical closure and is an essential part of what dualists contend. Hence, if we are actually reluctant to admit dualism and challenge the principle of physical closure, we are forced to limit the liberality of the principle of property causation by the strong supervenience constraint.

This argument is flawed, though. Firstly, once we acknowledge (following the principle of property causation) that there are causal relations between facts, we are bound to redefine our conception of the

closure of the physical world. The claim that all events are physical, and therefore have a complete physical explanation, becomes irrelevant. For we are concerned with not the causal connections between events as such, but between events in virtue of having certain properties. And these properties can certainly be physical, but also biological, geological, mental, and the like. Furthermore, the stronger contention that all facts are physical facts, that is, that all causal connections between events hold in virtue of having the physical properties they have, clearly goes beyond the reach of the principle of physical closure. In fact, it seems that the most that this principle could commit us to saying is that all physical properties of the world constitute a closed system, that is, that there is no physical property of an event which cannot be physically explained. But the principle of property causation is entirely compatible with such a principle even if the strong supervenience constraint is dropped out. For, on this interpretation, the principle of physical closure exclusively affects physical facts and its validity is independent of allowing for biological or mental facts not having a physical explanation.

Somebody might object, however, that, even if not all facts are physical facts, the principle of physical closure not only affects physical facts, but should also include the claim that all facts (physical and non-physical) must have a physical explanation. Let us refer to this new interpretation of the principle of physical closure as the *the principle of physical explanation*. We shall seek to show, though, that the introduction of this last principle either begs the question or falls short of what is required to support the intrinsicness condition.

To begin with, we should distinguish between a *strong* and a *weak* understanding of the principle of physical explanation. Specifically, the weak interpretation is associated with the global supervenience constraint. It trivially follows from this constraint that, for any fact of the form 'x is in the psychological state P,' there must be a physical supervenient base, that is, a set of physical truths about the world from which the truth of the statement that describes the psychological fact follows. What the weak principle of physical causation comes to saying is that a Laplacean demon, adequately endowed with the ability to produce psychological attributions, would be able to derive the truth of a psychological statement from his knowledge of the relevant set of

physical facts. It is crucial to stress, however, that, in allowing for the possibility of such a Laplacean demon, we are not assuming the existence of a systematic connection between the physical supervenient bases of mental facts that, by contrast, are systematically interlocked from an intentional perspective. We are not claiming that our Laplacean demon could, for instance, derive the physical supervenience base of 'John wants to drink a beer' from a suitable combination of the physical supervenient bases of properties like 'being John,' 'wanting,' 'drinking,' and 'being a beer.' Similarly, the weak principle of physical causation does not entail the existence of a systematic connection among the physical supervenient bases of psychological facts like 'John wants to drink a beer,' 'Peter wants to drink a cup of tea,' 'Peter believes that there is a beer,' etc., which, obviously, any intentional psychology should interrelate. The vindication of such a systematic connection is specific, by contrast, to the strong reading of the principle of physical explanation. Correspondingly, only the adoption of this strong principle could guarantee the existence of a systematic connection between physical and nonphysical explanations.

Going back to the argument for the intrinsicness* condition, we can easily see that the intrinsicness* condition can be substantiated by the strong principle of physical explanation. For it seems clear that only those properties of a system that are strongly supervenient on the physical properties of the system (even if relative to a context) can meet such a strong principle. Yet, in the context of our discussion, the strong principle of physical explanation cannot be used to ground either the intrinsicness* condition or the strong supervenient constraint without begging the question. For this strong principle is not stronger, but equivalent to the strong supervenience constraint we were supposed to substantiate.

At this point, it is crucial to notice that what we have called the strong supervenience constraint admits of a soft and a hard reading. On its soft reading, Kim is right when he claims that global supervenience entails strong supervenience³⁰. If two physically identical worlds share all their mental properties, we can hardly deny that there must be, for each particular instantiation of any given mental property, a sub-set of physical properties of these worlds that constitutes its supervenient base in both worlds. Needless to say, we have no problem

in accepting that a sub-set of the physical properties of a particular world defines the supervenient base of my present desire to drink water. But if this is all what follows from the soft understanding of the strong supervenience constraint, this cannot be the reading which lies behind the issue about mental causation, since mental contents would trivially satisfy such a modest constraint. Hence, it seems that the strong supervenience constraint deserves a more demanding interpretation, if it has to put a problematic constraint upon the causal efficacy of mental properties. We can, then, say the strong principle of physical explanation simply renders explicit the sort of requirement that, in the issue about mental causation, the strong supervenience constraint was meant to impose. Finally, we should say the weak principle of physical explanation falls obviously short of what is required to back up the intrinsicness* condition. Since, once the systematicity condition is abandoned, the strong supervenience constraint, which the intrinsicness* condition presupposes, stops being a necessary condition for causal efficacy.

Some people, though, may find this weak principle of physical explanation insufficient, that it does not square with our fundamental intuitions. For these people may conceive of causal chains between facts that are not nomically, systematically, dependent upon physical properties as utterly weird. The causal efficacy of such causal chains appearing as something mysterious. What is, however, the source of such uneasiness?. It seems clear that it cannot be strictly epistemic. Nobody doubts that we have the epistemic ability to detect what appear to be causal chains between facts, regardless of our epistemic ability to detect the supervenient base of them. Explanations and predictions in commonsense psychology manifestly rest upon such an ability. What is called into question, however, is whether we are dealing, in folk psychology, with *actual* causal chains and not merely chains that *appear to be* causal. Nevertheless, this is no longer an epistemic issue, but a metaphysical one. What we really want to decide is whether a certain apparent causal chain has a strong physical supervenient base, so that it could be recognized as an actual causal chain. Hence, the source of the uneasiness some people may feel is that, unless causal chains between nonphysical facts are strongly supervenient upon a physical base, we will not be able to say that they are actually efficacious. In

other words, what these people urge is that the strong supervenience constraint needs to be introduced in order to account for the distinction between real causal chains and mere regularities. This argument, though, has the structure of an inference to the best explanation, and what we propose is precisely an alternative story.

Why not consider the weak principle of physical explanation as the only metaphysical necessary condition, which, together with some additional epistemic requirements, will allow us to tell real from merely apparent causal chains? At first sight, some people may still feel dissatisfied and be inclined to say that this weak principle is insufficient, that some stronger metaphysical condition is required. Yet, our venture is that, at this point, these people may be begging the question: Can they show that our story is insufficient without previously assuming the strong principle of physical explanation and, therefore, the strong supervenience constraint? Unless they could do this, they would be begging the question. Of course, we have not shown that such an independent motivation could not be provided. Our argument in this paper has been notoriously more modest. We have confined ourselves to rebutting a classical argument for such a constraint, namely, the argument based on the rejection of dualism and the endorsement of the principle of physical closure. For our account acknowledges the closure of the physical facts of the world and avoids dualism by accepting that, for any particular causal chain, there must be a physical supervenient base. The supervenient base that would allow our Laplacean demon to impute contentful descriptions out of the perception of its physical base; but, as we have seen, this not enough to sustain the strong supervenience constraint.

We can, then, conclude that the rejection of dualism and the vindication of the principle of physical closure cannot be used to narrow down the scope of the principle of property causation by the strong supervenience constraint. Since dualism can be avoided and physical closure respected without this constraint being imposed. Furthermore, we are inclined to think (and intend to prove in a later paper) that the strong supervenience constraint is not required to account for any of our fundamental metaphysical intuitions, so that, unless we adopt it as primitive, we are bound to abandon it as metaphysically superfluous. From the abandonment of the strong supervenience constraint impor-

tant consequences follow for the intrinsicness condition and Dretske's naturalization program.

III (iii) Once the Strong Supervenience Constraint is Abandoned

If the strong supervenience constraint fails, the intrinsicness* condition becomes metaphysically unmotivated and Dretske's naturalization program comes out metaphysically superfluous. But let us go step by step.

As we have seen, Dretske views the intrinsicness condition as a necessary requirement for a fact to count as a triggering cause. Yet, in section III (i), we showed that the intrinsicness* condition rests upon the strong supervenience constraint. Hence, if this last constraint loses its foot, so does the intrinsicness* condition. The intrinsicness* condition was, however, the only metaphysical obstacle to be removed, in Dretske's approach, before recognizing extrinsic (*qua* extrinsic) properties as triggering causes. It trivially follows from this conclusion that the extrinsicness of mental contents does not prevent them from acting as triggering causes.

Correspondingly, Fregean thoughts will be able to operate as triggering causes, even if the reference of propositional attitudes are extrinsic properties of an organism. In other words, although the first part of Putnam's argument is correct: the reference of the propositional content of mental states is a relational or extrinsic property of an organism. The conclusion that Fodor and Dretske have drawn does not follow: Fregean thoughts cannot be triggering causes of our outputs and, therefore, cannot form a part of a scientific psychology which seeks to provide triggering explanations. We have argued, by contrast, that extrinsic properties of an organism can count as triggering causes of its outputs, so that a relevant part of Frege's intuitions about the identity of mental contents and their causal efficacy can be preserved.

Let us turn now to Dretske's program to naturalize mental content. Dretske's program becomes metaphysically irrelevant, once the strong supervenience constraint is abandoned. For Dretske's program is simply an attempt to show that mental contents meet the strong supervenience constraint as a necessary step to acknowledge the causal efficacy of mental content *qua* mental content. Metaphysical irrelevance

does not amount to epistemic irrelevance. In fact, Dretske's naturalizing attempt might be construed as promoting a fairly valuable epistemic goal, namely, increasing the degree of cohesion among the properties that the different kinds of theories that compose the scientific landscape deal with. In any event, our intuition is that, despite this epistemic reinterpretation, Dretske's naturalizing attempt fails, that is, does not manage to indicate how mental contents meet the strong supervenience constraint. In our opinion, Dretske's program is subject to the traditional objections to logical behaviourism. For Dretske's account relies on the behaviourist notion of discriminatory learning, and it has been soundly argued that the response to the question 'What is learned?' cannot be derived from the answer to the naturalistic question 'What is rewarded?', if vicious circularity is to be avoided³¹. But, of course, we have no room to discuss this point here.

IV CONCLUSIONS

This paper is an attempt to pursue Dretske's challenge to the intrinsicness condition. Dretske argues that extrinsic (*qua* extrinsic) properties can be structuring causes of a process, whereby they can participate in structuring explanations. He considers, though, that triggering causes must respect the intrinsicness condition, so that extrinsic (*qua* extrinsic) properties cannot enter triggering explanations. We have sought to go beyond Dretske's stance and defend the view that there is no metaphysical motivation to maintain the intrinsicness condition even for triggering causes (i.e., the intrinsicness* condition).

We have stressed that strong materialism entails the principle of property causation. For strong materialism presupposes the strong supervenience constraint, and this constraint can only be stated if one assumes that events can be causally efficacious *in virtue* of having certain properties, and this is what the principle of property causation urges.

Once we accept the principle of property causation, the intrinsicness* condition stops being trivially true and needs to be grounded upon the strong supervenience constraint. We have sought to show, though, that the adoption of the strong supervenience constraint is metaphysically unmotivated. For the principle of property causation al-

lows us to dispense with the strong supervenience constraint in the materialist attempt to escape dualism and vindicate the principle of physical closure. Hence, unless we are prepared to impose arbitrary metaphysical conditions or some further motivation is provided, we are bound to abandon the intrinsicness* condition. Consequently, mental contents *qua* mental contents can no longer be ruled out as triggering causes by the mere fact that they are extrinsic properties³² and, therefore, the fact that Fregean thoughts include, as Putnam pointed out, extrinsic properties does not prevent them from entering triggering explanations.³³

NOTES

(1) Cf. [Frege, 1966, p. 346, 360-2].

(2) For the sake of brevity, we will save the clause '*qua mental content*' in referring to the causal efficacy of mental content *qua* mental content, except where relevant to avoid misunderstandings.

(3) [Dretske, 1988, p. x]

(4) [Dretske, 1988, p. 79-80].

(5) This is, of course, a central issue in cognitive science. In the recent literature, J. Fodor [Fodor, 1987, p. 135; 1985, p. 78] maintains the strongest realist stance with respect to mental contents and their causal efficacy, although R.G. Millikan [Millikan, 1984, p. 8] and Dretske [Dretske, 1981, 1987] are also committed to intentional realism. By contrast, we find in Dennett [Dennett, 1981, 1987c, 1987d] a more instrumentalist position, whereas P.M. and P.S. Churchland

[Churchland 1989, 1986] vindicates eliminative materialism. For a map of the most relevant stances with regard to this issue: Fodor [Fodor, 1985], Dennett [Dennett, 1969, chapter 2, 1987a, 1987b, 1991] and also Lyons [Lyons, 1990a, 1990b].

(6) Cf. [Dretske, 1981, p. xi], and [Dretske, 1988, p. 80].

(7) Cf., for instance, [Kim, 1984, pp. 163-7].

(8) We can certainly distinguish between strong and weak supervenience. While strong supervenience is concerned with instantiation conditions that are valid across all possible worlds, weak supervenience only vindicates its validity relative to a given set of possible worlds. In any event, this distinction will not be explored, since it is not necessary to our purposes.

(9) As in Kim [Kim, 1984, p. 165-6], no commitment to a particular notion of possibility or necessity is required to introduce the distinction.

(10) Physical properties are the properties posited by the theories of physics, while nonphysical properties are the rest of the properties of the world.

(11) More sophisticated arguments in favour of a limited version of the intrinsicness condition will be deployed in due course. This remark is only meant to provide an initial and very rough motivation for the condition in question.

(12) Dretske defines 'extrinsic property' as follows: "The basic idea of an extrinsic property is that of a property a thing possesses, not because of the way *it* is, but because of the way *other* things, things to which it stands in certain relations, are (or were)" [Dretske, 1992, p. 3].

(13) For the classical case against the notion of narrow content, cf. Putnam [Putnam, 1975] and Burge [Burge, 1979]. A classical defence of this notion can be found in Fodor [Fodor, 1975], Fodor [Fodor, 1980] and Field [Field, 1978]. For interesting discussion of this issue cf. Woodfield [Woodfield, 1982] and Pettit and McDowell [Pettit and McDowell, 1986]. A more recent criticism of narrow content in Putnam [Putnam, 1988, chapters 1-3] and a revision of the concept in Fodor [Fodor, 1987, ch. 2], and Fodor [Fodor, 1991].

(14) Of course, these brief indications do not aim to elucidate Frege's notion of thought, but to mention a few aspects of what in the debate at stake have been described as Fregean thoughts.

(15) Cf. Fodor [Fodor, 1987, p. 46-53], where he comes to vindicate the existence of narrow contents even if they are essentially inexpressible.

(16) [Dretske, 1992, p. 6]

(17) [Dretske, 1990, p. 783].

(18) It may be advisable, at this juncture, to introduce a couple of qualifications. Firstly, not all behaviours include some sort of movement, even if they typically do [Dretske, 1987, p. 28]. Secondly, we must admit that there are processes that, like breathing and growing up, fit into Dretske's definition and yet we wouldn't be initially inclined to recognize them as behaviour. Nevertheless, Dretske is not worried by these apparent counterexamples because choosing to describe a process as behaviour is, in the end, interest-relative [Dretske, 1990] p. 784].

(19) Cf. [Dretske, 1987, p. 32].

(20) Cf. [Dretske, 1987, p. 18 and 36]. This distinction between process and output allows Dretske, as we shall see, to discriminate between the subject-matter of psychology and that of neuroscience (Cf. [Dretske, 1987, pp. 51-2]).

(21) [Dretske, forthcoming, sec. 1 par. 1].

(22) Cf. [Dretske, 1987, p. 42-3], and [Dretske, forthcoming, sec. 1, par. 4].

(23) Cf. [Dretske, 1987, p. 42-3], and [Dretske, forthcoming], sec. 1, par. 4].

(24) Cf. [Dretske, 1992, p. 8]. Similar arguments can be developed with biological examples where the same type of phenomenon may be due to two distinct evolutionary stories and, therefore, to two distinct structuring causes.

(25) Cf. [Fodor, 1987, 46-53], and [Kim, 1991, p. 67].

(26) [Dretske, 1991, p. 198].

(27) Cf. [Horgan, 1991, pp. 85-6].

(28) The principle of property causation enables us to discriminate between physiological and psychological explanations. Although both kinds of explanations deal with the same events, the principle in question allows us to say that they are concerned with distinct classes of facts. Dretske himself subscribes to this thesis and discriminates between outputs as the explananda of physiology and behaviours as the explananda of psychology.

(29) Cf. Kim [1984, 1990, 1989, 1991].

(30) Cf. [Kim, 1984, pp. 168-9].

(31) For classical objections to behaviourism, cf. [Chisholm, 1957 chapter 2]; and [Taylor, 1964, part II].

(32) Of course, further objections can be raised against the causal efficacy of mental contents but we have no room to handle them in this paper.

(33) We would like to thank Juan J. Acero, Fred Dretske, Manuel García-Carpintero, Tobies Grimaltos, Carlos J. Moya, David Owens, Lino San Juan, Barry Smith and Sergio Véliz for helpful remarks and discussion on an earlier version of this paper. Research for this paper has been funded by the Spanish Government's DGICYT as part of the projects PS-93-0178 and PB93-1049-C03-02. Our thanks to this institution for its generous help and encouragement.

REFERENCES

- [Burge, 1979] T. Burge, Individualism and the Mental, in: P. A. French et al. (eds.), *Midwest Studies in Philosophy* 4, 73-121.
- [Chisholm, 1957] R. Chisholm, *Perceiving*, Ithaca, Cornell University Press.
- [Churchland, 1986] P. S. Churchland, *Neurophilosophy: Toward a Unified Theory of Mind/Brain*, Cambridge, Mass.: The MIT Press.
- [Churchland, 1989] P. M. Churchland, *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, Cambridge, Mass.: The MIT Press.

- [Dennett, 1969] D. Dennett, *Content and Consciousness*, London: Routledge and Kegan Paul.
- [Dennett, 1981] D. Dennett, *Intentional Systems*, in: Dennett, D.: 1981, *Brainstorms*, Cambridge, Mass.: The MIT Press, 3-22.
- [Dennett, 1987a] D. Dennett, *Three Kinds of Intentional Psychology*, in: [Dennett, 1987e, 43-68].
- [Dennett, 1987b] D. Dennett, *Reflections: Instrumentalism Reconsidered*, in: [Dennett, 1987e, 69-81].
- [Dennett, 1987c] D. Dennett, *True Believers*, in: [Dennett, 1987e, 13-36].
- [Dennett, 1987d] D. Dennett, *Reflections: Real Patterns, Deeper Facts, and Empty Questions*, in: [Dennett, 1987e, 37-42].
- [Dennett, 1987e] D. Dennett, *The Intentional Stance*, Cambridge, Mass.: The MIT Press.
- [Dennett, 1991] D. Dennett, *Real Patterns*, *The Journal of Philosophy* 88, 27-51.
- [Dretske, 1981] F. Dretske, *Knowledge and the Flow of Information*, Oxford, Basil Blackwell.
- [Dretske, 1988] F. Dretske, *Explaining Behaviour. Reasons in a World of Causes*, Cambridge, Mass., The MIT Press.
- [Dretske, 1990] F. Dretske, *Précis of Explaining Behaviour: Reasons in a World of Causes*, in: *Philosophy and Phenomenological Research*, 50, 783-6.
- [Dretske, 1991] F. Dretske, *Dretske's Replies*, in: Brian P. McLaughlin, *Dretske and His Critics*, Oxford, Basil Blackwell, 180-221.
- [Dretske, 1992] F. Dretske, *What isn't wrong with Folk Psychology*, in: *MetaPhilosophy*, 23, 1-13.

- [Dretske, forthcoming] Mental Events as Structuring Causes of Behaviour, in: Al Mele and John Heil (eds.), *Mental Causation*, Oxford, Oxford University Press.
- [Field, 1978] H. Field, *Mental Representation*, *Erkenntnis* 13, 9-61.
- [Fodor, 1975] J. Fodor, *The Language of Thought*, The Harvester Press, Hassocks, Sussex.
- [Fodor, 1980] J. Fodor, *Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology*, *Behavioural and Brain Sciences* 3, 63-110.
- [Fodor, 1985] J. Fodor, *Fodor's Guide to Mental Representation*, *Mind* 94, 79-100.
- [Fodor, 1987] J. Fodor, *Psychosemantics*, Cambridge, Mass.: The MIT Press.
- [Fodor, 1991] J. Fodor, *A Modal Argument For Narrow Content*, *The Journal of Philosophy* 83, 5-26.
- [Frege, 1966] G. Frege, *Logische Untersuchungen*, Göttingen and Zurich, Vandenhoeck and Ruprecht.
- [Horgan, 1991] T. Horgan, *Actions, Reasons, and the Explanatory Role of Content*, in: Brian P. McLaughlin, Dretske and His Critics, Oxford, Basil Blackwell, 73-101.
- [Kim, 1984] J. Kim, *Concepts of Supervenience*, in: *Philosophy and Phenomenological Research*, 45, 153-76.
- [Kim, 1990] J. Kim, *Explanatory Exclusion and the Problem of Mental Causation*, in: Villanueva, E. (ed.), *Information, Semantics, and Epistemology*, Oxford, Basil Blackwell, 36-55.

- [Kim, 1989] J. Kim, Mechanism, Purpose, and Explanatory Exclusion, in: Tomberlin, J. (ed.) *Philosophical Perspectives*, 3. Philosophy of Mind and Action Theory, Atascadero (California), Ridgeview Publishing Company, 77-108.
- [Kim, 1991] J. Kim, Dretske on How Reasons Explain Behaviour, in: Brian P. McLaughlin, *Dretske and His Critics*, Oxford, Basil Blackwell, 52-72.
- [Lyons, 1990a] W. Lyons, Intentionality and Modern Philosophical Psychology, I: The Modern Reduction of Intentionality, *Philosophical Psychology* 3, 247-69.
- [Lyons, 1990b] W. Lyons, Intentionality and Modern Philosophical Psychology, II: The Return to Representation, *Philosophical Psychology* 4, 83-102.
- [Millikan, 1984] R. G. Millikan, *Language, Thought, and Other Biological Categories*, Cambridge, Mass.: The MIT Press.
- [McDowell, 1986] P. Pettit and J. McDowell (eds.), *Subject, Thought and Context*, Oxford University Press, Oxford.
- [Putnam, 1975] H. Putnam, The Meaning and 'Meaning,' in: Putnam, H.: 1975, *Philosophical Papers*, v. 2: *Mind, Language and Reality*, Cambridge University Press, Cambridge, 215-271.
- [Putnam, 1988] H. Putnam, *Representation and Reality*, Cambridge Mass., The MIT Press.
- [Taylor, 1964] C. Taylor, *The Explanation of Behaviour*, London, Routledge and Kegan Paul.
- [Woodfield, 1982] A. Woodfield (ed), *Thought and Object. Essays on Intentionality*, Clarendon Press, Oxford.

KARLOVY VARY **STUDIES**

in Reference and Meaning



*Roberto Casati, Josep Corbí, Steven Davis,
Manuel García-Carpintero, Olav Gjelsvik,
Anthony Grayling, Petr Kořátko, Diego Marconi,
Pavel Materna, Nenad Mišćević, Alex Orenstein,
Josep Prades, Greg Ray, Marietje van der Schaar,
Stephen Schiffer, Petr Sgall, Gianfranco Soldati,
Charles Travis, Achille Varzi, Alberto Voltolini,
Takashi Yagisawa, Eddy Zemach*

*Edited by
James Hill & Petr Kořátko*

Copyright © 1995 The authors. The author(s) given at the head of each contribution is(are) the holder(s) of the copyright, except in those cases where it is specifically stated otherwise in the text.

All right reserved.

No part of this publication may be reproduced or distributed in any form or by any means, or stored in a data base or retrieval system, without the prior written permission of the holder(s) of copyright.

Published by FILOSOFIA-ΦΙΛΟΣΟΦΙΑ Publications
of the Institute of Philosophy of the Academy of Sciences
of the Czech Republic

Edited by James Hill & Petr Kořátko

Cover © by Vladimír Havlík

Design and typeset by Martin Pokorný

Printed by Ústav jaderných informací, Zbraslav, a.s. (Prague)

ISBN 80-7007-077-3

Contents

Preface	7
Stephen Schiffer, Meanings and their Nature	9
Greg Ray, Untokened Sentences in Actual Languages	29
Anthony C. Grayling, Perfect Speaker Theory	43
Steven Davis, The Grice Program and Expression Meaning	60
Eddy M. Zemach, Seeing Zeus	68
Roberto Casati and Gianfranco Soldati, On the Perception of Abstract Objects	89
Manuel García-Carpintero, Dretske on the Nature and Explanatory Role of Meaning	114
Josep E. Corbí and Josep L. Prades, Mental Contents in a World of Causes	148
Petr Sgall, From Meaning Via Reference to Content	172
Petr Kořátko, Definite Descriptions: The Nicest Vehicle of Reference I Know	184
Takashi Yagisawa, Reference <i>ex Machina</i>	215
Diego Marconi, Dual Aspect Theories and Reference	243
Alberto Voltolini, Indexinames	258
Marietje van der Schaar, The Cognitive Value of Indexical Sentences: Kaplan versus Husserl	286
Alex Orenstein, How to Get Something from Nothing	300