

(This is a pre-print text)

Mind, Brain, and Downward Causation*

Carlos J. Moya. University of Valencia (carlos.moya@uv.es)

This paper has three parts. In the first, we present some basic notions, such as mental and intentional causation, and causation itself; we then proceed to state some theses, inspired by the natural sciences and to show how they seem to conflict with the acceptance of intentional causation. This conflict is the core of the problem of mental causation. In the second part, we present and discuss briefly some proposals to deal with this problem and try to bring to light their difficulties. And, in the third part, we put forward some tentative suggestions as to how the problem of intentional causation might eventually find a solution.

1. Mental and intentional causation: the problem

Mental causation is the process by which certain events or states, *by virtue of being mental, i.e. of having mental properties*, give rise to changes in the physical properties of the world, by giving rise to physical events or states such as brain changes, muscles' contractions and overt bodily actions. The caveat "by virtue of having mental properties" is intended to exclude cases in which some events that actually have mental properties cause physical events, but where those mental properties are completely irrelevant to the causal relation. These would be cases of physical, not mental causation. Dretske makes this point very clearly. He writes:

Something possessing content, or having content, can *be* a cause without its possessing that content or having that meaning being at all relevant to its causal powers. A soprano's upper-register supplications may shatter glass, but their meaning is irrelevant to their having this effect. Their effect on the glass would be

* This paper was presented, in an abbreviated form, in the STOQ Project Workshop "Methodology of Neuroscience. Between Philosophical, Phenomenological and Empirical Approaches", which took place at the Pontifical University of the Holy Cross, Rome, June 7-8, 2010. I am grateful to Prof. Juan José Sanguinetti for inviting me to take part in this event, as well as to the rest of participants for their comments and criticisms. This paper has been elaborated in the context of the research project FFI2009-09686, awarded by the Spanish Ministry of Science and Innovation, to which thanks are due.

the same if they meant nothing at all or something entirely different (Dretske 1988, p. 79).

Intuitively, the glass' shattering as a result of the singing is a case of purely physical, not mental, causation, even if the singing has mental properties.

A second point of clarification concerns the notion of a mental property. With Dretske and many other thinkers, we assume that semantic properties such as content or meaning are among mental properties. Having content or meaning is a sufficient condition for some state or event to be mental. In this sense, a meaningful speech act counts as mental, as do beliefs, purposes, hopes or desires. However, having content or meaning is not necessary for mentality, for pure sensations ('raw feels', 'qualia') are intuitively mental events or states even if they lack any content or meaning. Interesting as these "qualitative" properties of some mental states may be, in this paper we shall restrict our attention to semantic properties and states or events having them. Given this restriction, we may afford to remain neutral for what respects other possible properties of those states characteristically emphasized by Cartesianism, such as privacy, subjectivity or first-person infallibility or incorrigibility. So, those who conceive of meaning and content as social and publicly accessible properties do still face the problem of mental causation, provided that they take states and events with those properties to be mental.

Let us now say some words about causation. This is in itself a hard and controversial philosophical issue. However, a minimal assumption about causation seems to be part and parcel of any reasonable view of it, namely that causes make a difference to their effects in that at least they make the occurrence of the latter more probable than it would have been without them. Raising the probability of some event's occurrence seems to be a necessary condition for something to be a cause of that event. It is hard to see how event A could possibly be a cause of event B if B's occurrence would have been equally probable in the absence of A. Applied to the causal relevance of properties, this assumption becomes something like this: in order for property P of event A to be causally relevant for A's causing event B, then, all else being equal, A's having P should make B's occurrence more probable than it would be otherwise. This assumption is very weak and very plausible as well; and it is compatible with different views of causality, deterministic or probabilistic, nomological, counterfactual, and others. Deterministic views, for example, may interpret the assumption so that, all else

being equal, the presence of the cause raises the probability of the effect to 1. In the above quotation, Dretske employs a version of this assumption in order to show the causal irrelevance of the meaning of the soprano's singing for the glass' shattering; the singing's meaning does not make the probability of the shattering higher than it would be without it, for the shattering would have taken place exactly as it did even if the singing had borne no meaning at all; therefore, meaning was causally irrelevant to the causal relation at hand. Now, we think that accepting this minimal assumption about causality is enough for the problem of mental causation to arise. We will try to justify this contention below.

Mental causation is a special case of the more general phenomenon of mind/body causal interaction, namely mind-to-body causation. As we anticipated, in this paper we will focus mainly on a particular species of mental causation, which may be called intentional causation. In intentional causation, the mental properties of the causing events or states are semantic properties, especially meanings and contents. We are all too familiar with intentional causation. Our presence in this workshop is a result of a process of intentional causation in the indicated sense, provided that we have come to Rome by virtue of having certain intentional states, purposes, beliefs, preferences and the like.

Some philosophers have thought, and some still think,¹ that, even if intentional states or events sometimes explain an agent's action by giving her reasons for performing it, this sort of explanation is not causal and those intentional states (reasons) are not causes of the action. Now, it may seem that, in assuming that there is a problem of intentional causation, we are begging the question against these non-causalist thinkers, for they would deny that there is such a thing as intentional causation and so any problem about it at all. However, if we are right that accepting the above assumption about causation is enough for the problem of mental causation to arise, then mental causation, with a different name, is also a problem for non-causalists. In effect, if the above assumption about causation is correct, then, by accepting the truth of statements of the form "S A-ed for reason R", they are implicitly accepting that S's having R made S's A-ing more likely than it would have been otherwise. If they did not accept this, then it is hard to see how the explanation could be true. So, non-causalists seem also committed, concerning reasons, to the indicated assumption about causes and

¹ Among the former, L. Wittgenstein, G. E. M. Anscombe, W. Dray, P. Winch, etc. Among the latter, C. Ginet, J. Tanney, D. Hutto.

causally relevant properties; and, it being so, they must admit that reasons, understood as mental states or events, make a difference to the physical world and it is legitimate to ask them for an explanation of that fact.² Under a different label, then, mental causation is also a problem for non-causalists about reasons and reasons-explanations.

In spite of its central importance in human life and of our apparent familiarity with it, intentional causation seems to conflict with some general views inspired in natural sciences, such as: 1) the causal closure of the physical domain; 2) a layered and hierarchical view of the world, in which the physical level is metaphysically basic; and 3) the view that the brain, which controls our intentional behaviour, is a purely syntactic, not a semantic device, so that it cannot be sensitive to contents or meanings.

According to 1), in Baker's words, "every physical property-instantiation that has a cause at t has a complete physical cause at t " (Baker 1993, p. 78). Roughly, every physical event has a complete physical cause. The physical domain, then, is causally closed: there is no need to get out of it in order to find complete, fully sufficient causes, and corresponding causal explanations, of what happens within it. Causal closure does not apply to orders of phenomena other than the physical one. Some biological phenomena may have chemical or microphysical causes. And the same holds for psychological phenomena: certain psychological changes may be caused by chemical changes in the blood or the brain. This is the point of calling certain chemical substances "psychoactive". And there is no need to go that far: the process of forming beliefs based on perception of our surroundings shows the causal openness of the psychological realm. Now, at least apparently, intentional causation seems to conflict with the causal closure of the physical domain. If sometimes we act intentionally because of what we believe, desire, or intend, then, since our so acting involves several physical changes, these changes also occur, at least partly, because of the contents of those states. And this looks like a breach of the causal closure of the physical. In fact, it is not even required that we actually perform those actions; it is enough, for that breach, that the probability of our so acting, and of the corresponding physical changes, is raised by the contents of our beliefs or intentions. That some changes in the probability degrees of certain physical changes are due to mental events *as mental* looks like an intrusion of non-physical factors into the physical domain and so as a violation of its

² In accepting the indicated assumption, however, they are not ipso facto accepting that reasons are causes, for, as we pointed out, the assumption is only necessary, but not sufficient, for an event to be a cause or for a property to be causally relevant to a causal relation.

causal closure. But why should this violation be a problem? Jaegwon Kim answers as follows: “[T]o give up this [physical closure] principle is to acknowledge that there can in principle be no complete physical theory of physical phenomena, that theoretical physics, insofar as it aspires to be a complete theory, must cease to be pure physics and invoke irreducibly non-physical causal powers” (Kim 1993, p. 209).

Let us move on to 2). This is a metaphysical worldview inspired in natural science. According to it, the world has a layered and hierarchical structure of ontological levels or strata, some of which are more basic than others, so that more basic levels can exist without the less basic ones, but not vice versa. Roughly, the level of physical entities and properties would be the more basic, followed by the chemical, biological and psychological levels in this order. To this ontological hierarchy corresponds an epistemological and explanatory hierarchy of sciences: physics, chemistry, biology and the different sciences of life (including neuroscience), and, finally, psychology and the human and social sciences. This hierarchy rests on an order that is both chronological and causal: the physicochemical stratum existed long before, and gave rise to, the biological one, which, in turn, came into existence before, and gave rise to, the mental stratum. According to this hierarchical view of the world, there are causal relations between different levels, but causation always proceeds in an upward direction. Events and processes of lower strata affect causally those of higher strata, but are not affected by them. We may characterize the relations among the levels in terms of supervenience, understood as a relation of metaphysical dependence. Entities, properties, and causal processes of a given level supervene on those of lower levels, which are in turn their supervenience bases. In this context, the physical level is the most basic of all: their properties and causal processes do not supervene on those of any other level, and these in turn supervene ultimately on the physical level. Now, intentional causation seems to violate these hierarchical and supervenience relations, for it implies that certain mental events, by virtue of their mental properties, bring about changes at lower levels, including the most basic microphysical entities and processes. In this respect, intentional causation seems to involve an objectionable and mysterious form of downward causation.

Concerning 3), a position that might be called “syntacticalism”,³ the claim is that brain processes and structures can only be sensitive to physical or, at most, syntactical

³ The view is associated with Stich (1983) and is also related to what Fodor calls “the Formality Constraint” on naturalistic theories of mind.

properties of signs or events, not to their semantic properties, such as content or meaning. Syntacticalism is closely related to computational functionalism, for which computers are models of the relations between mind and body. On this perspective, mental processes should be conceived as operations on signs, or inputs generally, according to their form, not to their meaning. But the deep questions raised by syntacticalism do not depend on accepting computational functionalism. In Daniel Dennett's words, the view would be as follows;

[T]he brain is first and foremost a syntactic engine, which can be fruitfully viewed as reliably mimicking a semantic engine, but in which meanings themselves never overrule, overpower, or so much as influence the brute mechanistic or syntactic flow of local causation in the nervous system. (A semantic engine, I claim, is a mechanistic impossibility – like a perpetual motion machine...) (Dennett 1991, p. 119).

Now, if the brain is the system that controls and brings about voluntary behaviour, then only physical and syntactical properties will be causally relevant to such behaviour.⁴ Concerning it, then, semantic properties appear to be causally idle, purely epiphenomenal properties. So far, intentional causation, which implies the causal relevance of content and meaning, would seem to be incompatible with syntacticalism.

The philosophical problem of intentional causation, and mental causation generally, is partly the problem of what to do with these conflicts. Our firm conviction about intentional causation is that it is sometimes what we believe, desire, or intend, i.e., the content of these states or events (as well as the fact that we believe, desire, or intend that content to be true), a decisive factor in leading us to act in certain ways. In other words, we are convinced that content is causally relevant to our intentional behaviour. On the other hand, the general views inspired in natural sciences seem also to be important and well motivated. A good solution to the problem of intentional causation, then, would be to show how to reconcile that firm everyday conviction, which is part and parcel of our self-image as rational agents, with the scientifically inspired views. However, showing this consistency has proved to be quite a hard enterprise.

⁴ As Kim expresses this idea: "Syntacticalism appears to entail that intentional properties of mental states, those properties in which their mentality consists, are causally irrelevant" (Kim 1991, p. 55).

2. Some proposals about intentional causation

In fact, as far as we know, existing proposals to deal with this problem are less than satisfactory. We shall present, and comment on, some of them.

One possibility, associated with some non-causalists and some of Wittgenstein's followers,⁵ is to hold that reasons, understood as mental states with content, explain intentional actions, not bodily movements, whereas brain states and events explain bodily movements, not intentional actions. This "dual explanandum strategy", as it might be called, is intended to show that there is no real conflict or tension between causal explanations in natural sciences and everyday reasons explanations, for their respective explananda are not the same. As I see things, however, the strategy has problems. Even if it is true that mental states with content (reasons) explain behaviour viewed or described as intentional, such mental states are also causally responsible, on each occasion in which an agent acts intentionally, for the occurrence of bodily movements involved in that intentional action, for presumably these movements would not have occurred if the agent had not performed the intentional action at issue; and, at least in some cases, she would not have performed it unless she had had those mental states with the contents they had. The fact that we do not have a unified theory that connects concepts or descriptions of mental states with concepts or descriptions of bodily movements, understood as physical and physiological events, does not imply that the former are not causally responsible for the latter. And so the problem of how this intentional, mental-to-physical, downward causation can take place does not vanish.

Another possibility is reductionism, in its many varieties. Think of so-called property- or type-identity theory. According to it, mental properties just are (in the sense of "are identical with") neurophysiologic properties, and so ultimately physical properties. So, mental properties are causally relevant because physical properties are causally relevant and mental properties are physical properties. Intentional causation involves no mystery. Mental causation does not breach the principle of the causal closure of the physical domain, for it is physical causation itself. Nor does it involve downward causation. In the context of the hierarchical worldview, the idea would be that the psychological entities and properties do not constitute a level on its own, but are part of the biological level; there is, then, no downward causal influence from the

⁵ But not only them. Peacocke also seems to endorse such a strategy: "It is true that brain states and efferent connections will be enough to explain any given bodily movement. But what is distinctively explained by a set of externalist states are relational facts about events or objects, relational facts which go beyond mere bodily movements" (Peacocke 1993, p. 207)

mental/psychological to the biological level; and, at the end, biological properties are, or supervene on, physicochemical properties; instantiations of biological properties do not have autonomous causal powers, but inherit them from those instantiations of physicochemical properties which they supervene on (or are identical with). So, causation is always upwards, and downward causation is a mere appearance. It may be objected that type-identity theory does not respect the conviction that content is causally relevant in getting us to act. But the theory can account for this conviction as well, since contents themselves are, on this view, neurophysiologic properties, and so as causally relevant as these can be.

The main problem with type-identity theory is that the prospects for its truth are rather bleak. Initially, the first proponents of this theory restricted it to sensations and other “qualitative” states, and for good reasons, since the theory is more plausible concerning these states, especially in the case of quite primitive sensations, such as pain, than about intentional states. A necessary condition for intentional properties to be identical with neurophysiologic properties is that there be strict nomological correlations between them, so that, e.g., the brains of all human subjects who believe that the Earth moves around the Sun share one and the same neurophysiologic property. But there are good empirical reasons for thinking that there are no such strict correlations. It is very plausible to think that, if intentional properties, like the indicated belief, have neurophysiologic realizations (or supervenience bases) at all, these realizations or bases can be very different in different subjects, depending on a variety of factors such as their genetic endowment, mother language, personal development, and others. They can even be different in the same individual at different times. Multiple realizability of mental properties is, then, a serious objection against type-identity theory.

Still more damaging for this theory is semantic and content externalism, according to which meanings and intentional contents do not even supervene on neural properties, so that two physically identical individuals can differ in the meaning of their utterances and in the content of their beliefs and other intentional states owing to differences in their respective environments, natural or social, even if these differences have no internal reflection in their brains and bodies. Hilary Putnam (1975), Tyler Burge (1979) and other philosophers have forcefully argued for semantic and content externalism.

But if content externalism stands in the way of neurophysiologic reductionism, it is also an obstacle to intentional causation itself, for it would seem that properties that are causally relevant to behaviour generation should be internal to an agent's body, whereas, if externalism is true, meaning and content are extrinsic and relational properties, not fixed solely by factors internal to the brain or body. This favours syntacticalism, the view that only signs, not their meanings or contents, can be causally relevant to behaviour production.

Other reductive proposals, such as Kim's, also face difficulties. Type identity theory rested on a Nagelian model of reduction, on which reduction involves derivation of the laws of the reduced theory from the reducing one, with the help of biconditionals ("bridge laws") that connect the predicates of both theories, thus giving rise to one-to-one equivalence relation between properties. Kim sees this model of reduction as seriously defective (cf. Kim 2000, pp. 90 ff.) He advocates a different model, which involves two main steps. First, a functional analysis of the property to be reduced, in which that property (or the concept thereof) is conceived just in terms of its typical causal relations to other properties, states and phenomena, not in terms of its physical constitution. This makes this functionally analysed property into a second-order property, which quantifies existentially over first-order properties. The second step is the statement of a relation of realization between the second-order property and the first-order properties that meet the causal specifications given in the functional analysis. The first-order properties realize the abstract, second-order functional property. This model allows for reconciliation between reduction and multiple realization. Fragility, for instance, would be a second-order property, realized by several different molecular structures. Mental properties should also be conceived as second-order, functional properties, which are realized by several distinct physical, presumably neural, properties. According to Kim, realization implies supervenience: if physical property P realizes mental property M, then M supervenes on P. Concerning the causal powers of mental events and properties, he holds that they are identical with the causal powers of the neural events and properties which realize them and on which they supervene:

On a reductionist position of this sort ... the causal powers of mental properties turn out to be just those of their physical realizers, and there are no new causal powers brought into the world by mental properties (Kim 2000, p. 118)

So, on Kim's proposal, mental events or states cause other events or states by virtue of the physical properties and causal powers of their neural realizers or supervenience bases, not by virtue of their own mental properties. Since (unlike what happens in identity theory and at least in cases of multiple realizers) mental properties are not identical with the neural properties that realize them, mental properties appear to be causally idle. In this respect, Kim's functional reductionist model does not fare essentially better, for what concerns the threat of epiphenomenalism about mental properties, than his earlier model of mental causation as supervenient causation. So, in the context of the latter, he writes:

[A]lthough content properties are relational ... the causal power of a given instance of a content property [e.g. a particular belief, C. M.] lies wholly in the causal power of the neural state on which it supervenes (Kim 1991, p. 67)

So far, rather than accounting for mental causation, Kim's reductionism seems actually to rule it out.

Beyond this central problem, Kim's proposal faces other difficulties as well. First, content externalism, if true, excludes not only identity of mental and neural properties, but also supervenience of the former on the latter. And, given that realization implies supervenience, excluding supervenience implies excluding realization as well. As far as we see, without neural realization and supervenience of the mental, however, Kim's proposal collapses. Second, relations between mental states, and between them and intentional behaviour, show certain semantic and inferential patterns without which human beings could not be held to be rational beings. However, since mental states and intentional behaviour are supervenient on neural bases, which relate to each other in a purely causal and factual way, according to certain natural laws, those rational patterns appear as a sort of brute fact or surprising coincidence, and are left unexplained within Kim's model.

Dennett (1991) intends to show that, though the real causal work is done by physical properties of the brain, not by meanings or contents, there is a reliable harmony between physical properties and semantic properties, a harmony established by evolution and conditioning processes. Dennett's proposal could be of help with our last objection to Kim's position. Nevertheless, on Dennett's perspective, no less than on Kim's, intentional causation becomes a mere appearance. Our strong conviction that

sometimes we act in certain ways because of what we believe or want, or because of the meaning of others' (and our own) utterances, would be a mere illusion, brought about by 'harmonizing' processes. On Dennett's proposal, we mimic semantic systems; we act *as if* we were sensitive to meanings, but we are not and cannot be so. As in Kim's case, this is not to explain intentional causation, but actually to deny it.

As for non-reductive materialist proposals, it is widely acknowledged that one of them, namely anomalous monism, does not accord mental properties enough causal relevance. Concerning emergentism, it would seem to have problems with some of the views inspired by science, such as the causal closure of the physical domain and the layered, hierarchical worldview. But new and increasingly refined versions of emergentism are being designed and we should not discard that some of them might be able to overcome such problems.⁶

Dretske's interesting attempt to naturalize meaning and content (cf. Dretske 1988, 2009) goes a certain distance towards reconciling intentional causation with scientifically inspired views. According to Dretske, mental causes are what he calls "structuring causes" of behaviour. Behaviour itself "is not the bodily movements that internal events cause, but the causing of these movements by internal events" (Dretske 2009, p. 21). Bodily movements are caused and explained by the physical properties of internal events, but in order to understand why certain internal events cause those movements we have to advert to semantic properties of the internal events; certain internal events are recruited as causes of certain bodily movements because of what they indicate about the environment, and not only because of their intrinsic physical properties. There is still some distance between the notion of indication and that of meaning or content, a distance that has to do with the intensionality and the possibility of error (e.g., false beliefs) which characterize intentional states; this is an endemic problem for informational and naturalizing theories of content and meaning, such as Dretske's, Fodor's, or Millikan's; but the proposal is suggestive and it might be further refined so as to solve this problem; in fact, some aspects of those proposals, such as Millikan's notion of proper function, are quite plausible in this respect. Another problem of Dretske's position is that intentional states seem to figure, in ordinary reasons explanation, not (or not just) as structuring causes but, in Dretske's terms, as

⁶ Interesting insights can be found in Murphy (2010).

triggering causes as well.⁷ We usually do not want to know why a certain desire led someone to act in a certain way, but why she acted that way, and we try to find some beliefs or desires of her in order to answer this question; and these beliefs and desires would seem to function as ordinary, triggering causal factors.⁸

Robert Van Gulick (1993) has also put forward a valuable proposal for understanding downward causation that has some points of similarity with Dretske's. According to Van Gulick, the predicates of special sciences individuate patterns of organization of physical constituents, and these patterns have causal powers of their own, which cannot be understood simply in terms of those of their ultimate physical constituents and the laws of physics (cf. Van Gulick, p. 250). He writes:

[H]igher-order patterns can have a degree of independence from their underlying physical realizations and can exert what might be called downward causal influences without requiring any objectionable form of emergentism by which higher-order properties would alter the underlying laws of physics. Higher-order properties act by the *selective activation* of physical powers not by their *alteration* (Van Gulick 1993, p. 252)

Van Gulick's proposal applies especially well to cases in which, say, biological patterns activate underlying physical powers, but we still need to understand how this process functions in the case of intentional patterns. But the approach is undoubtedly promising.⁹

3. Some suggestions about intentional causation

It would be unbearably pretentious to think that we can provide a solution to the problem of intentional causation. But here are at least some tentative suggestions as to how we could begin to understand the causal relevance of content and meaning to brain

⁷ As Horgan writes: "[W]e think of the reason, *qua* reason, as relevant not merely (and not mainly) insofar as its content might figure in some past-operative structuring cause of the present action, but rather as a full-fledged *triggering* cause of the action, and indeed as a *current* triggering cause. But Dretske is committed to denying that reasons, described as reasons, have this kind of here-and-now explanatory relevance" (Horgan 1991, pp. 88-89).

⁸ In relation to this, Dretske's view of behaviour as the process by which certain internal events cause bodily movements looks quite strained and alien to our ordinary concept.

⁹ Murphy's paper refers quite extensively to Van Gulick's ideas as congenial to her own emergentist view. For a more detailed exposition of Dretske's and Van Gulick's proposals cf. my 2004, chapter 14.

processes and overt behaviour, and so downward causation, without breaking a reasonable allegiance to science.¹⁰

A step towards that end is a reflection on the process through which human beings are introduced into social semantic systems, such as musical notation and its interpretation with one's voice or a musical instrument, or just, to take a more common example, ordinary language. Each sound or utterance made by the learner requires internal neurophysiologic causal antecedents; in this respect, we have upward causation; increasingly better performance as a musical interpreter or as a speaker requires, it seems, increasingly complex neural connections, which causally sustain such performances; this is again upward, physical causation; however, which neural connections are selected among the myriad that are possible depends upon the objective semantic content of the musical or linguistic signs; connections that give rise to correct sounds or utterances are reinforced during the learning process, while those that produce the wrong results are not; and the difference is determined by the semantic rules that govern the systems; in this sense, certain connections within the brain, as well as with nerves and muscles, are selected and shaped through a process of downward causation: from the contents and meanings of the musical and linguistic signs, according to the semantic rules of musical notation and ordinary language, to neural and neurophysiologic connections. To this extent, content and meaning, which, as externalism has it, go beyond the individual's brain and bear an objectivity of their own, are causally responsible for the actual shaping of the neural connections and networks required for a competent musical or linguistic performance.

As we see things, an important source of resistance against the reality of intentional causation is a certain view of human beings that might be taken to be, somewhat paradoxically, a sort of materialistic dualism. Whereas, according to old forms of dualism, a human being is essentially a soul or thinking thing that contingently inhabits a body, for this new dualism a human being is essentially a brain that contingently inhabits (the rest of) the body. In current philosophy of mind, and even epistemology, the importance of the brain, which we do not want to deny, has been magnified, with a corresponding devaluation of other parts of the human body, such as the tongue or the hands, which tend to appear as mere peripheral appendices at the service of the brain, and ultimately dispensable; every kind of bodily activity, such as

¹⁰ We have come to see these suggestions, in retrospect, as having some debt to Dretske's and Van Gulick's views.

the limbs' movements, is always conceived as a mere effect of brain processes, and never as a source or cause of what happens or exists within the brain. This conception makes it unintelligible how properties belonging to higher ontological levels, and especially meanings or contents, might exert any influence on the structure and processes of the brain. It is also an important source of syntacticalism. However, on the basis of our preceding considerations, we can see that the agents' hands, their tongues, eyes and vocal muscles, play a decisive and non-dispensable role in the progressive shaping of neural connections: it is only through these overt physical movements and acts that the process of education and correction, and the corresponding processes of selection and reinforcement of neural networks, can take place; only in this way can the learner become a competent interpreter and speaker. An isolated, disembodied brain could not, by itself, acquire the internal complex shaping that supports musical or linguistic performance; it is by selecting certain overt movements among those emitted by the agent that the right neural connections are selected and reinforced; and so, though each such movement is preceded and caused by a neurological event, there is also a causal feedback that goes from objective contents or meanings, through certain physical movements of the limbs, selected according to those objective semantic properties, to the inner structure of the brain and its internal connections. A disembodied brain, without the help of the limbs and other parts of the body, could not learn meanings and contents and, if thinking requires them, could not *think* either. Subjects of thought, of beliefs, desires, purposes, and meanings are human beings, not brains by themselves.

We may grant Dennett that brains are syntactic systems. But a living human being is not only a brain, and she is sensitive to meaning. In fact, in the process of learning, say, to interpret musical notation, masters have to be sensitive to the sounds that correspond to the musical signs and to the relation between signs and sounds, which are the signs' meanings; otherwise they would not be able to correct the learners' performance. And learners become progressively sensitive to such relations, and so to meanings, as their competence increases. And this sensitivity helps explain that human subjects can respond to signs by virtue of their meaning, not just their form or syntax. Meaning is thereby causally relevant to behaviour production. Once a fairly complex network of neural connections has been established on the basis of socialization and learning, human beings can understand and respond to the meaning of words and other signs emitted either by others or by themselves, even in a sub vocalized, purely 'inner' way. And meanings become progressively 'transparent' to us: our experience in our

mother language is that we ‘hear’ meanings directly, with signs receding to the background.

Far from being alien to scientific enlightenment, accepting the reality of mental and intentional downward causation seems to be required in order to understand and explain certain important phenomena. Think for example, in connection with our previous remarks about music, of the amazing synchronization of movements by the violinists of a good orchestra. There must be important common patterns in their neural networks that account for that synchrony. And these common patterns would not be there if processes of downward causation, starting from the objective meanings of musical signs, had not affected the brains, nerves, and muscles of each of those subjects while they learned to read musical notation and to interpret it with their violins. A “brute mechanistic or syntactic flow of local causation in the nervous system”, to use Dennett’s words, a process of purely upward causation, unaffected by meanings and contents, could only have given rise to that impressive coordination of movements by miracle, by virtue of coincidences of an astronomically low degree of probability. So, it is partly the search for a rational explanation of social and cultural communities and communal phenomena that should lead us to accept the influence of meanings and contents upon neurobiological processes in the brain and try to understand how it is possible without entering the realm of obscurity or mystery.

REFERENCES

Baker, Lynne R. (1993), “Metaphysics and Mental Causation”, in Heil & Mele (eds.), pp. 75-95.

Burge, Tyler (1979), “Individualism and the Mental”, in P. A. French et al. (eds.), *Midwest Studies in Philosophy*, 4, pp. 73-121.

— (1982) “Other Bodies”, in A. Woodfield (ed.), *Thought and Object. Essays on Intentionality*, Oxford, Clarendon Press, pp. 97-120.

Dennett, Daniel (1991), “Ways of Establishing Harmony”, in McLaughlin (ed.), pp. 118-130.

Dretske, Frederick (1988), *Explaining Behavior: Reasons in a World of Causes*, Cambridge MA, MIT Press.

— (2009), “What Must Actions Be for Reasons to Explain Them?”, in C. Sandis (ed.), *New Essays on the Explanation of Action*, New York, Palgrave Macmillan, pp. 13-21.

Heil, John and Mele, Alfred (eds.) (1993), *Mental Causation*, Oxford, Clarendon Press.

Horgan, Terence (1991), “Actions, Reasons, and the Explanatory Role of Content”, in McLaughlin (ed.), pp. 73-101.

Kim, Jaegwon (1991), “Dretske on How Reasons Explain Behavior”, in McLaughlin (ed.), pp. 52-72.

— (1993), “The Non-Reductivist’s Troubles with Mental Causation”, in Heil & Mele (eds.), pp. 189-210.

— (2000), *Mind in a Physical World*, Cambridge MA, MIT Press.

McLaughlin, Brian P. (ed.) (1991), *Dretske and His Critics*, Oxford, Basil Blackwell.

Moya, Carlos J. (2004), *Filosofía de la mente*, Valencia, Publicacions de la Universitat de València.

Murphy, Nancey (2010), “Avoiding Neurobiological Reductionism: The Role of Downward Causation in Complex Systems”, manuscript written for the Workshop “Methodology of Neuroscience”, University of the Holy Cross, Rome, June 7-8, 2010.

Peacocke, Christopher (1993), “Externalist Explanation”, *Proceedings of the Aristotelian Society* 93, pp. 203-230.

Putnam, Hilary (1975), “The Meaning of ‘Meaning’”, in *Philosophical Papers, Vol. 2: Mind, Language, and Reality*, Cambridge, Cambridge University Press, pp. 215-271.

Stich, Stephen (1983), *From Folk Psychology to Cognitive Science*, Cambridge MA, MIT Press.

Van Gulick, Robert (1993), "Who's in Charge Here? And Who's doing All the Work?", in Heil and Mele (eds.), pp. 233-256.