

Chapter 10. Connectionism and causality

“Willing” is not the name of an action; and so not the name of any voluntary action either. And my use of a wrong expression came from our wanting to think of willing as an immediate non-causal bringing about. (PI 613)

The twin issues of connectionism and causality are closely linked in that if connectionism is a correct understanding of the way that the mind works then that has significant implications for the kind of account of causality that would be appropriate in psychological explanation.

I. The requirements of causal realism.

John Stuart Mill carefully laid out a flatfooted materialist stance towards action that has set the current agenda for those interested in the topic.

To my apprehension a volition is not an efficient but simply a physical cause. Our will causes our bodily actions in the same sense, and in no others, in which cold causes ice, or a spark causes an explosion of gunpowder. The volition, a state of our mind, is the antecedent; the motion of our limbs in conformity to the volition is the consequent. (System of Logic, 355)

The causal thesis that Mill initiated can summarise, at least in its present form, as follows:

An action = a trying + the bodily movement caused by it

According to Mill, the kind of causation involved in action explanation is material causation and does not require any of the more contentious claims that might arise were we to appeal to formal, efficient, or final causes. These other types of cause are not usually regarded as properly part of naturalistic explanations as they involve the kinds of explanatory move that may not easily be mapped onto the workings of a mechanism such as the human body or brain is often construed to be.

A formal cause or explanation of a phenomenon mentions only the reason why the phenomenon has the form that it does. For instance one might ask, Why is that figure the shape it is? One might answer “Because it is a triangle and to be a triangle it must be a closed figure bounded by three straight sides.” This explanation specifies the form that the phenomenon must have if it is to count as a thing of the kind in question but, quite evidently, it does not explain the genesis of the phenomenon in any other way. In action explanation this causal explanation would do no more than say something of the type “he moved like that because he was trying to swat the fly”.

Often such a formal explanation goes hand in hand with a teleological or “final” causal explanation in that it tells us the intention that guided the genesis of the phenomenon but, since Hume, we have been suspicious of this kind of explanation of action in that it seems to inculcate something mental or ideal the mechanism of action of which in bringing about a physical event such as a bodily movement is obscure.

The final cause of an event is usually taken as the end or goal at which the causal process terminates and although it answers a question as to why the event occurred as it did it is usually assimilated in naturalistic philosophy of mind to the events which comprise the effective reasons which lead to the action (where these are then construed as mental states and events preceding the bodily movements constituting the action).

The category of efficient causation as the originating process in the genesis of an event is also contentious if it is taken to mean anything other than the material cause. This is

because the naturalistic origins for events are usually sought in the causal chain of physical events antecedent to the action. A material cause is the one implicated in this kind of explanation and it is the physical state giving rise to the event.

Thus we are left with a Millian insistence on the idea that until we have an efficient or material causal explanation in terms of antecedent physical events, we do not have a respectably naturalistic explanation of human action.

Kim is quite forthright about this claim and espouses a kind of causal realism about mental life.

According to causal realism ... causal connections hold independently of anyone's intentional states – in particular epistemological or doxastic states ... The realist believes ... that causal relations – the same ones – would hold even if there were no conscious beings to contemplate them or reasons concerning them. ... according to causal realism every event has a unique and determinate causal history whose character is entirely independent of our representation of it. (1993, p235)

Kim's outline of causal realism, when applied to mental causation, generates a further condition on such an explanation: "representation independence". This condition entails that the cause of a mental event, such as an action, is independent of the way that the cause or the effect is represented. We must therefore explore these metaphysical requirements underpinning efficient material causes of the type required in naturalistic mental explanation.

There seem to be four major conceptual requirements which underpin a causal realist thesis of the type envisaged by Kim (Davidson, 1980; Morris, 1986):

C(i) if A caused B then A is an event which has an existence in the realm of nature independent of B;

C(ii) it is genuinely contingent that B occurs in conditions where A has occurred;

C(iii) there is a connection between A and B which can be conceptually assimilated to our conceptions of causal connection;

C(iv) the cause and effect (and their relation to one another) exist and operate independently of the way that anyone, including the agent thinks of them.

These features are easily justified in relation to well grounded claims about material causes and all four are related in ways that can be clarified by one or two examples.

However, I will argue that these four conditions, taken together, cannot be satisfied by anything which is a plausible candidate for a reference of naturalistic terms constituting descriptions of the functions of mind and brain.

So, to the well grounded claims. Take the spurious claim that a drug, *sleepazide*, causes a person to sleep because of its "dormative potency". Let us say that if one asked about the existence of the "dormative potency" then we should respond "Well that is just its capacity to put people to sleep". Now here there is either a non-claim masquerading as an explanation or an admission that, although one does not know what it is, there is some independently existing property of *sleepazide* that induces sleep in human beings. If there is no such conception as the second on the agenda then there is no genuine causal explanation.

Imagine, as a second case, that one tried to argue that the legs of a particular table were the length they were because the top was at a given height. One might object that this is not a causal claim in any but a formal sense – both facts are a logical part of the fact that the table has certain dimensions. But the causal claim is made good when it is pointed out that "It ain't necessarily so". Say that the table top was required to be at that height so that the queen of the household could see what was on the table for dinner from her favourite armchair. Then there would be a causal factor relevant to the genesis of the particular leg length which did operate via the height of the table top. Admittedly this is

part of an efficient causal story (in Aristotle's sense) but nevertheless it seems to be the kind of thing that is relevant to the production of an effect according to a naturalistic account.

A third case is the claim by a fellow of Magdalen College, Oxford that the Magdalen bells cause the Christchurch bells to ring. Now if this is a genuine causal claim, then it is not a bare constant conjunction claim (a la Hume), therefore the most that the claim would provoke is amusement at the irrational chauvinism of Oxford fellows. But what if the explanation was then given that the Christchurch clock was broken and that the bell ringers at Christchurch assembled promptly on the quarter hour to ring the Christchurch bells the moment they heard the Magdalen bells ringing (or in this day and age there was a powerful directional microphone tripping a switch to set the Christchurch bells ringing); in the latter cases the claim has credence because we have the right kind of connection.

Lastly, and very relevant to the present case is the claim that fire is caused by the oxidation and release of heat by combustible substances. The claim illustrates one or two important points. One ought to note that the alleged cause is not specified independent of the effect ("combustible substances"). This is not, however, metaphysically important because it merely serves as an epistemic marker for a dispositional characteristic uniting a diverse set of substances. The nature of that characteristic can be independently investigated by physico-chemical science and described independent of the shared disposition. The status of the causal claim is, however, contentious on the grounds that the alleged cause just is just part of the complex event we call fire and therefore is not an independent and antecedent condition producing the effect. We could say that the two components of the claim are just different descriptions of the same state of affairs each of which has its own cognitive significance to us.

It is instructive to distil these metaphysical requirements so as to see how mental explanation might conform to a naturalistic account embedding material causation as its active principle.

II. Causal realism and mental explanation.

To give a physicalist (material causation based) account of mental explanation we would need:

- (i) a chain of individuable events $C_1 - C_n$ which provided a clear mechanism of production for an action A^* ;
- (ii) a physical description of the event which constitutes an action (e.g. A^*) even if our normal epistemic access to that event was not in terms of that description;
- (iii) a nesting of the principle of connection between $C_1 - C_n$ and A^* in a set of natural laws which are based on plausible causal chains.

(This set of conditions is closely related to the metaphysical claims C(i) to C(iv) listed as essential to the thesis of causal realism above.

A physical account of the type envisaged is powerfully motivated by the need to forge a principled distinction between the reasons for which an agent acted and reasons held by the agent which would make sense of the behaviour concerned. That we need such an account to explain mental activity is a *sine qua non* for Kim: "To deny the mental causal powers would seem tantamount to denting it the ability to explain anything or to make a difference to anything; to be real and to have the power to enter into causal relations arguably go hand in hand, at least for concrete entities" (1993, p121). Kim proposes that there is a supervenience relation between the mental and physical features of any event with mental properties or a mental description and that therefore one can find a causal

chain which includes the mental antecedents of an action such as A* and A* itself. This entails that “a psychological explanation of a given event can be construed as supervenient and hence dependent on its physical explanations”(1993, p128). The kind of explanation involved in Kim’s construal is summarized quite nicely by Childs in his discussion of causality and interpretation.

Elements of any psychological story can be correlated, one-to-one, with the elements of an internal physical story about what is going on in a subject’s brain and nervous system. So distinctions, for example between acting for one reason and acting for another, or between really and only apparently remembering, correspond straightforwardly to distinctions between different physical aetiologies that a subject’s behaviour, or behavioural disposition, might have. (1994, p179-80)

Dretske (1988, p43), suggests that the causal account is based on the implicitly two component explanation of how a cognitive system has got structured in a certain kind of way so that a characteristic input tends to produce a certain kind of output and a specification of a suitable input which triggers it to produce the pattern constitutive of an action. We could summarise Dretske’s analysis in terms of structuring cause (c.f. the presence of combustibles in the right condition) plus triggering cause (spark) where, for mental explanation, both are conceived as causally realised representations with proper representational relations to sets of conditions in the real world (1993, p147). Dretske invokes indicators functions such that “an element becomes an F representation (representing things as F) if it has the function of indicating that they are F.”(p148). This picture is reductive in so far as it presupposes nothing about the world which cannot be governed by naturalistic laws. The functions discussed here are the kinds of normal or proper functions discussed by Millikan as generated and sustained by naturalistic processes which have played a certain role in the evolution of the function as a cognitive function with a semantic content (“an information supplying function”[p149]). On this account the explanation of physical movements by mental content is not problematic as, in Dretske’s terms “Content is already in the physical domain”(p149). We ought however to note that the accounts of explanation within this genre are aimed at physical movements and not actions *per se* except in so far as they can be identified with those movements.

A general objection undermining such accounts would arise if we argued that the model of events upon which it depends was misconceived and that the type of explanation involved in action was not physical causal explanation but more akin to the naturalistic explanations of ecology. This would directly parallel the points made about genetic adaptation and cultural adaptation and would advance the case of a Wittgensteinian construal of intention and action. I will argue that Wittgenstein holds a picture at odds with a Millian style of causal thesis as espoused by Kim and the causal realists. It remains to be seen whether anything that one might try to put in place of that theory will sit comfortably with the data of neuroscience so as to satisfy the following desiderata:

- D1 it fits with the data relating to our natural history;
- D2 it preserves the intuitions left intact by a Wittgensteinian critique of mental realism;
- D3 it assimilates mental explanation to a philosophically respectable scheme of naturalistic explanation.

I will in fact claim that a notion of something like an efficient cause (the originating intentional act) which incorporates a telos (a purpose or end at which the act is aimed) gives us a better understanding of mental causation and one which cannot ground a reductive deterministic view but which will mesh very well with the account of social causation to follow.

It is commonly held that a thoroughgoing acceptance of what I have called the Aristotelian claim – that mind or *psyche* is a set of functions of the human body – is not only compatible with but even required by and explicated by the Millian analysis and therefore that such a view is incompatible with any robust doctrine of freedom of the will. The argument usually goes through the twin premises that all mental events, including actions or intentions are physical events and that physical events form a more basic level of explanation than mental events. The latter claim is thought to imply that physical or brain events cause their mental counterparts (as Mill argues). I think that this belief is false and the argument fallacious. To defend this claim I will argue that the kind of explanation required to do the work of mental explanation or the explanation of behaviour in terms of thoughts, attitudes, beliefs, desires, and so on is a type of explanation that is not conceptualizable in terms of strict deterministic causality but rather embeds an assumption that the psychological subject is a free agent who creatively conceives of certain ends on the basis of skills he or she has mastered. If I can make a case for that view then certain claims about the causal priority of brain events over intentions and actions cannot be sustained.

III. Worries about causal realism

The most flatfooted approach to the requirement for causal realism about mental explanation is a version of identity theory but the position I have argued for whereby there is no type-type correlation between mental events and brain events makes such a position untenable. The possibility of a kind of antirealism about physical event types in the brain based on Wittgensteinian and connectionist construals of the basis of mental activity also threatens any robust functionalist reading and I have also argued that LOT and related theories do not seem plausible given a thorough neurophilosophical analysis.

Childs has offered a fairly sweeping rejection of claims of this type. Childs argues on three bases for the conclusion that even if “the causal explanations of common-sense psychology are physically based, ... they presuppose no detailed psychophysical correlations”(p179).

He notes that the lack of strict psychophysical laws gives us a principled reason to reject any kind of type-identity theory which implies that in regard to lawlike physical causal relations there is little prospect of fulfilling Kim’s requirements.;

He remarks on the fact that the norms of rationality are not easily codifiable (a fact for which I have argued elsewhere [Gillett 1999, p]). This lack of codifiability defeats any lawlike functionalism based on propositional attitude psychology, in fact, this is almost a corollary of the relation of classical computationalism.

The sort of general isomorphism required by the common-sense functionalist’s project, between psychological causal roles and physically specifiable causal roles, is incompatible with the uncodifiability of rationality.(p80)

Childs also gives us reason to suspect token identity theory and the relatively weak supervenience claims that follow from it. he argues that the only way a token identity could, in principle, be conceived would be to identify a unique causal role filled both by a particular mental entity M’ and a physical entity P’. But this story is able to be iterated and therefore we must be able to defend exactly what the uncodifiability defeats – a substantial mapping of mental states (connected by norms of rationality construed according to our holistic understanding of situated human beings) on to the structure of physical or brain states connected by exceptionless physical or syntactic laws.

Childs also reports a further argument produced by Hornsby (1981) which dovetails rather nicely with the connectionist considerations to which I am about to appeal.

Hornsby's argument concerns the impossibility of individuating a neurophysiological event which can be distilled from the complex and ramifying neural activity going on at any moment to identify with any given mental event such as M'. Now, if there is no suitable *relatum* on one side of a relation, then there is no relation. I shall offer further support for the view that we do not find a ready taxonomy of physically identifiable events with which to correlate the elements in a mental explanation as we consider what actually does happen in the brain.

To some this may seem to leave us with a kind of explanatory nihilism but that conclusion only follows from what we might call "physicalist fundamentalism", the view that the only real explanations that can be given of any phenomena are explanations which make use of terms properly employed in the natural sciences. This, as Stich has noted in relation to certain of its variants, is a particularly difficult view to state in any coherent or convincing way.

For instance we could try to claim that the only real things are things that appear in well-formulated natural laws. These are things that can be individuated (described and quantifiable over) in terms given to us by the physical sciences and one might want to claim that apart from such things that nothing else can "make a difference to anything"(Kim, 1995, p121). But this claim is hard to defend in that *The al Qaeda organization*, *Amnesty International*, *racism in the Balkans*, and the *International Monetary Fund* do seem to be real enough and to make a real difference to the world but they are clearly not physical things in the sense described. However leaving these odd individuals apart for the moment we can, as Stich does, appeal to more everyday things like revolvers and candlesticks. Even though these things do not appear in physical laws or quantification sets, it is clear that they have causal effects, one can be killed by either of them (and not just in a board game). What is more the fact that they can kill is explicable in terms of natural laws. But racism in the Balkans has also killed a lot of people as has The al Qaeda organization. Undaunted, our physical fundamentalist may still argue that any real causal effects that these things have or had are only to be explained by their genuine scientifically describable properties. However these things are not individuable in those terms. What is more, any attempt to classify history, politics and economics as sciences and the individuals they allow us to identify as natural kinds is quite contentious. Philosophers of natural science such as Hull would not allow such a move and usually the physical fundamentalists want something more because they are not content unless one says that the effects of any of these historical (or non-natural-kind) entities are only really explained and able to form part explananda as long as they are identified with states and objects apt to fill place-holding roles in the laws and descriptions of the natural sciences. But this is impossibly restrictive in terms of mental and social explanations and the kinds of phenomena they are indispensable in explaining and we are left with boundary disputes precisely in the areas in which neurophilosophy is interested.

This becomes apparent, for instance, in the type of genuine causal claim which contrasts with the sleepazide type of claim outlined above: the real reason that morphine relieves pain is that its molecular shape is suited to binding with endorphin receptors in the brain. Here a complex psychological property – relieving pain – is explained by an underlying chemical property and an associated biological property. The partial and parasitic nature of this claim in relation to an understanding of pain and the role it plays in human life and activity is obvious. The complexity of pain has already been discussed as has the fact that what an individual calls pain may be a state of suffering with very little relation to anything as biologically simple and well-characterised as "a critical level of firing in C fibres" or something similar. I have argued that to understand any psychological property of human beings requires a kind of holism that is exactly mirrored in what we now see

are adequate accounts of neurophysiological processing. But now consider the following two assertions:

1. Morphine at a dose of D' relieves pain depending on an individual's prior consumption and the pain level and type involved.
2. A causal link is only a true causal link if the entities referred to are physically described.

We have discovered truth 1 in the clinical practice surrounding pain management and it is located somewhere in the overlap of domains occupied by sciences like pharmacology, biochemistry, and neurobiology. We have a hope of understanding the basic features of such an explanation by attending to empirical work in these domains but, as should be obvious, that may still rely on extensive work dealing with the emergent and irreducible property that pain is coming to be seen as.

When it comes to claim 2 (A causal link is only a true causal link if the entities referred to are physically described), the case is different. There is no science which has discovered this truth or told us what it would be for this claim to be true. This is a metaphysical claim (Kant would call it a synthetic a priori claim) and metaphysics does not (except in scientific circles) conduct itself in terms dictated by natural science. In fact it is far more plausible to believe that explanation is a way of discerning relationships between events described in various ways such that we can make intelligent attempts to intervene in those processes when and where we want to. This claim might however be thought to embody a kind of pragmatist scepticism (highly congenial to Wittgenstein's approach) about the metaphysical status of causality.

Let us say we accept this pragmatism and are prepared to live with the fact that mechanistic or functional type explanations as they figure in the syntactic theory of mind are not what mental explanation is all about. What then follows from this? Almost nothing of consequence except that we have to give up the idea that there might be machine-like or formally codifiable ways of framing adequate explanations of human behaviour. But that is hardly news as Davidson, Stich, and Childs (among many others) have argued for exactly that position. It may seem shocking if one is captivated (as Kim seems to be) by the naïve picture in which there is a bivalence between mechanism and inexplicability. But let us say that we do have to lift our eyes to the socio-cultural to get a good grip on human behaviour and that low level descriptions are always and only subservient to understanding how the well adapted personal and interpersonally-related agent work as he or she should, would that be so bad? Only for folks who like their philosophy and psychology to be mechanistic. What is surprising is that the idea that neuropsychology could so conduct itself was thrown into doubt many years ago and the writing was on the wall that we would have to look at emergent levels of explanation to do good neurophilosophy.

IV. The case of Luria

Luria, wiring almost thirty years ago, realised that human voluntary action is, in a very real sense, summative activity, an expression of the whole psychological system of a person working in concert to organise a strategy designed to bring about an envisaged and positively valued future state (1973, p246). This distinguished his understanding of action from the then dominant model in which one had a moderated reflex arc or conditioned (by past-events acting causally as stimuli) response to a pattern of input from the environment. That model has, for understandable reasons, remained the dominant paradigm because of its affinity with the idea of material causation as the mechanism of production of behaviour and other psychological events. The mechanistic model of biological activity was, ironically, bequeathed to us by Descartes and the Millian model of efficient material causation by antecedent causes is merely an

extension of that model.

However Luria notices several conceptual problems with this model.

First, there is the fact of *non-simplicity of relations*: “the intention or motor task ... is hardly ever a simple response to an external stimulus”(p248) because it is aimed at an effect which may have an idiosyncratic, remote, or very indirect relation to the conditions in which the action is exhibited depending on the significance of the state of affairs and the exhibited action in the psychological world of the individual concerned. (This, of course, does not look promising for law-like generalisations.)

Second, there is the problem of *physical non-stereotypy* in terms of the natural kinds of neuroscience. Luria points out that when we look at a piece of behaviour “the invariant motor task is fulfilled not by a constant, fixed set, but by a varying set of movements which, however, lead to the constant invariant effect”(1973, p248). What is more this “variation in the modes of performance of the movement is essential in principle for any active movement”(p249) because of its intentional character (it is directed or aimed at an actual environmental effect or target – as conceptualised in a certain way - not at the fulfilment of a plan based on a picture of the environment internal to the organism).

Third, there is a negotiated “comparison between the action as it is performed and the original intention”(p250) because both are (actual-) world-involving. Thus it is hard to speak of an executive output as if some cognitive message was sent to the body to be implemented or rendered into a pattern of movement. It is, however, just this sort of output that is the natural terminus of a Millian causal chain.

Lastly, Luria was conversant with the theory, later to be taken up by Dennett, in which action - and internal conscious activity in general – is linked closely to imperatives and the use of self-directed speech. For Luria this immediately invoked the social character of thought and action in which “the function previously shared between two persons became a method of organization of the higher forms of behaviour”(1973, p247). This is not a flatfooted doctrine of self-command but rather a sophisticated intimation of the role played by speech and shared signification within a human group in the elaboration of adaptive activity in a shared environment.

Luria goes on to discuss the role of the whole neuraxis in producing actions and intentions to act “carried out by an equally complex dynamic ‘constellation’ or concertedly working brain zones, each of which makes its own contribution (1973, p255). But notice that the only contribution which these diverse areas make is to tend towards the action A*. Thus there may be no independent specification of the relevant antecedents.

So we should return to the present chapter of the mental explanation story. I have reinstated a role for language as part of the representational structure we employ in cognition but not in a way which allows me to be realistic about the causal structure and syntax of proposition like elements in our cognitive machinery. The challenge is to flesh out what kinds of explanation I can offer for our actions. In particular I need to say what it is that allows actions to be endowed with the content ascriptions proper to action explanations rather than mere bodily response explanations.

V. Connectionism and explanation

The fundamental arguments that are levelled against connectionist accounts of mind, have to do, as we have already seen, with issues of compositionality and syntax. The classical computational theory of mind, represented by Fodor and supporting cast, holds that mental representations are actual explicit tokens of components of propositions functionally realised in the brain and that the syntax and grammatical rules of something like transformational grammar govern the transactions between them. This leads to a

theory which is somewhat misleadingly called folk- (or belief-desire) psychology according to which we account for people's actions by constructing syllogisms which look as if they work according to a quasi-logical set of operations. In fact the connectionist account only opposes this particular strong functionalist and realist theory that has appropriated the name folk psychology and not any approach to the mind that makes use of folk psychology terms in such a way as to offer true explanations but eschew the strong programme.

Macdonald summarises the connectionist position as follows:

1. it does not construe cognitive processes as involving symbol manipulation; and
2. the whole network is conceived of as a dynamic system which, once supplied with initial input to the input units, spreads inhibitions and excitations among its hidden units, eventually exciting certain output units and inhibiting others.(1995, 1p9)

Macdonald correctly identifies the need for the connectionists to show how the systematicity of connections which is a function of semantic composition might be produced if there is no explicit realisation of the relevant semantic components. This is highly relevant to any theory which attempts to configure action explanation so that it takes account of mental contents and the semantics thereof. Smolensky, as we have seen is somewhat persuaded by classical representational theory and responds in terms of a "conscious rule interpreter", conveniently forgetting the point that is pressed by Kant, Wittgenstein, and Millikan that there must be a way of obeying a rule which does not involve an interpretation. In order to preserve the intentionality of the content ascriptions needed for action explanation as a performance aimed at an environmental (or external) effect, Millikan also opts for "relational descriptions and relational explanations of function"(1995, 1p289). This is an anti-internalist position cashed out in terms of relations to objects and features in the actual world. Taken together these remarks suggest a certain reading of Smolensky's claim about the types of connectionist model that aim to give content-sensitive explanations of action.

The resulting connectionist model of mental processing is characterized by context-sensitive constituents, approximately (but not exactly) compositional semantics, massively parallel structure-sensitive processing, statistical inference and statistical learning with structured representations.(1995, 11p192)

Wittgenstein's pragmatic approach to mental content and Millikan's ecological constraints would gloss this summary claim such that connectionist systems realise patterns of excitation linked both to distributed patterns of neural activity tied to features and objects commonly encountered in the environment on the one hand and to semantic constituents on the other. I would link the latter to the signifiers applied to them in human discourse so that there would be a dynamic interplay between the level of compositional elements in natural languages and the level of conditions appearing in nature when certain signs are exhibited by language users. We could call such an orientation "cognitive interactionism" and thereby indicate that the "appearings" detected by creatures in the natural world are potentiated by their conceptual systems. This would not imply that the contents of those appearings could be metaphysically reduced to an internalist specification and is far removed from any kind of idealism about content, but it does acknowledge the interdependence of content fixing conditions and the epistemic system producing states and events which capture those conditions. This grounds a neuro-cognitive reflection of certain contextual conditions to be holistically mapped on to a range of combinations of semantic units such that a fairly accurate determination of the crucial conditions for the exhibition of a given semantic item could be distilled from the balance struck between multiply interacting patterns of excitation and inhibition. We can illustrate as follows.

Imagine I see a brown cow in one field and a black and white cow in another. The utterance that tends to be evoked (from self or others) by certain multi-track dispositions as a result of the first is 'brown cow in the field' and that evoked in the second is 'black and white cow in the field'. Now, given the points of congruence and distinction between these two utterances and some "good-enough" rules about where descriptors or property terms normally show up in English, I am enabled to discern correspondences which emerge between terms (signifiers) and what they signify. Compositionality arises from these good-enough correspondences, the multi-tracked dispositional links between the conditions in which utterances are produced on different occasions and the linguistic expressions typically exhibited in those situations. Such an account has the virtue of constructing language-components from a dynamic constitutional analysis of mental representation which does not assume that there are type-specifiable semantically apt functional components realised in brain processing units independent of exposure to speech communities.

This analysis also has, as I have noted, suggestive links to Frege's context principle and a constrained holism reminiscent of both Quine and Davidson. The cognitive negotiation that I have indicated takes into account the external conditions, the whole utterances produced, the behaviours exhibited and a (probably innate) set of assumptions (or framework of a less formal type) in which we recognise and assimilate ourselves to creatures like us. There is a huge body of empirical evidence for the existence of the latter framing assumptions in the infant mind.

This approach has interesting implications in the light of Andy Clark's conciliatory arguments for the claim that "any system complex enough to count as a believer will reveal (under some *post hoc* analysis) semantically clustered patterns of excitation" (1995.II, p352). The modest realist position gives rise, on his part, to an uncertainty about the proper response to the threat of eliminativism, especially if one concedes, as Wittgenstein and Dennett have hinted, that there might be no echo of the mental in the physical states described by neuroscience;

just *suppose* that mental processing turned out to be completely resistant to all attempts to find a well-motivated scientific description which follows the folk individuation of mental states. In such a case I am currently unsure what to say.(p354)

My suggestion is that situations as detected by the individual thinker are also penetrated by the words and utterances used in them. Thus I have argued that the thinker detects three kinds of input:

- (i) the conditions of utterance (both human and contextual);
- (ii) the parsed utterance; and
- (iii) an idea of the behavioural intentions of the utterer.

There then occurs in the distributed network a dynamic interplay between the excitation paths preferentially activated in distributed areas of the brain dealing with each of these three types of information. The result of the interplay is to assign signified conditions to components of the signifying expression and an actual-world content laden significance to the whole utterance where both of these achievements maximise their conformity to two different sets of constraints -- *first*, the semantic structure of the discourse currently framing the event and *second*, the ecologically salient objects of attention in the current situation. These two sets of constraints arise in two distinct kinds of ways. The first arises from discourse and the influences (largely interpersonal and cultural) to be discerned there and the second from the thinker's adaptation to the physical environment with its impingements on his or her body. Neglecting either set of constraints will cripple the semantic engine. We may now offer a tentative but highly plausible interpretation of a remark by Smolensky and go on to assess theories of

causality likely to yield dividends in psychological explanation.

From the point of view of the connectionist model builder, the class of networks that might model a cognitive agent who believes that dogs have fur is not a genuine kind at all but simply a chaotic disjunctive set. (1995.11p357)

Smolensky's scepticism (even nihilism) is predicated on an attitude prevailing in philosophy of mind where, following Mill, it is assumed that bottom-up explanation predominates over any idea that there can be genuine explanations arising from any other level particularly one where regularities and explanatory influences have no analogue at the physical level (the kinds of explanation and description that are found in discursive psychology for instance). The same source gives rise to Clark's puzzlement, Dretske's invocation of naturalisable indicator functions (1988), Millikan's biologism about cognitive content, and the debates raging between Smolensky and Fodor about whether semantic compositionality can be properly captured by a mathematical function defined in terms of physically describable properties of network activation.

On the view I have recommended, the drive to identify a physical basis for the individuation of mental types should not trouble us because we should recognise that salient cultural and social patterns will configure neural networks to realise their regularities in a way that is coeval to (or even more prominent than) the configuring influences found in the natural regularities of the unmodified environment. I have argued that human ecology is ecology in an artifactual world and human signifiers are themselves informationally promiscuous artifacts which need have no essential connection to their evolutionary forbears. Bearing in mind both the need for a sensible social naturalism and a rapprochement with connectionist approaches to cognitive neuroscience, we can begin to consider the kinds of causal or other explanatory relations that may be properly invoked in psychology and then progress to a comparison with Wittgenstein's remarks on intention and action.

VI. Anti-Causality and other types of explanation

Most of those who want to account for the phenomenon of mental explanation in naturalistic terms subscribe to the Millian view of explanation I have already introduced. As embraced by Kim, this claims that, in some sense, mental explanations must be dependent on or determined by physical properties on which the mental phenomena, in his view, supervene. Kim argues that any alternative views in which we posit equally adequate and self-standing mental and physico-causal explanations face us with a kind of "epistemic predicament" (1995, 1p126) in which we are uncertain as to the real explanation of an event. Kim, predictably, opts for a fairly strong kind of supervenience – "*ceteris paribus* supervenience" (p136) and is modest about the details. He proposes that only a metaphysical claim in which the dependence of the mental on the physical has "mental properties lawfully correlate, *ceteris paribus*, with their base physical properties" (p136) as the solution of Davidson's problem about the identification of the *reason for which* an act has been done (rather than merely a sufficient *reason for* relation, where the states appearing in the *reason for* have an appropriate content to explain the action and are plausibly attributable to the agent but are conceivably not the reason for which the action was performed). Kim suggests that the solution to the problem of individuating the actual reason for which the agent acted must be solved causally, thereby agreeing with Davidson that mental explanation is a species of causal explanation. His main argument for the conclusion that mental explanation must somehow supervene upon "fundamental physical processes" (1p138) is similar to Millikan's claim that "the only alternative to biological design, in our sense of 'design' is sheer coincidence, freak accident" (1p262).

I want to argue that Tyler Burge's externalist account of mental explanation can be

extended by appeal to the material above about second nature and the connectionist views that I have espoused to guide us to a real alternative to the stark choice proffered by Millikan and Kim.

Individualism and internalism jointly form Burge's target and are crucial components of the idea that actions as physical events are causally produced by a configuration of internal states, events and processes which capture or realise mental content. Burge pursues a powerful assumption grounding internalism, here in the philosophy of visual perception:

our concept of objectivity is such that no one objective entity that we visually represent is such that it must vary with or be typed so as necessarily to match exactly, an individual's proximal stimuli and discriminative abilities. (1995, p198)

We might render this more persuasive by arguing, with Millikan (but not in accordance with her flat-footed realism), that successful creatures would surely develop cognitive mechanisms that tracked or were linked to external or environmental objects and features rather than internal ones. The reason is obvious and closely linked to the fact that actions, as I have argued (with Luria) are intentional (in the phenomenological/externalist sense), in that they are essentially directed upon objects in the environment external to the brain. I have visited this argument in Chapter 1 in discussing the possibility of swampman but a reacquaintance in relation to the Burge-Fodor debate is in order. Burge's arguments are essentially twofold:

- (a) twin earth arguments;
- (b) arguments for the social or group determination of content.

The *twin earth arguments* are familiar to all involved in the debate between internalists and externalists about meaning and their import is well summarised in the following argument.

(a1) The primitives of a theory of intentionality are individuated by conditions external to the organism which figure as perceptual and behavioural targets for that organism.
(a2) If these conditions were different the content of the subject's mental states would be different.

(a3) It is conceivable that the proximal states and events of the subject would be indiscernibly similar.

(a4) In such a case the intentional properties of the agent would be different even though their internal states would be the same (as far as anyone could tell armed with the taxonomy of individualist physical theory). (Macdonald and Macdonald, 1995, p162).

En route to establishing this set of claims Burge also notes that our sensori-motor systems act as a whole directed through multiple modes at the objects we encounter and the events we find ourselves caught up in. In these interactions there are two relata and what passes between them is an essential part of explaining what is going on in the (or any) subjective relatum. For this reason we cannot individuate the intentional properties of the subjective relatum without invoking that (significantly external) understanding of what is happening.

(b) Burge briefly mentions the fact (stressed in the discussion of cultural evolution above) that most of the important tasks at which the human organism tries to succeed "arise through relation to the natural and social worlds" (1995, p201). Given that our relations to the natural world concludes give many philosophers reason to consider externalism persuasive, the mention of the social world would seem to clinch the case. The social world *par excellence* is cognitively tractable only on the basis that there is a locus of properties and independent changes in the environment which require us to posit an agent radically external to the subject doing the perceiving and "tracking" of the things which we also perceive and track (as the discussion of autism has repeatedly emphasised [Gillett, 1999]).

Burge concludes his discussion of these issues with the appealing claim that, [a]n illuminating philosophy of psychology must do justice ... to psychology's attempt to account for tasks that we succeed and fail at, *where these tasks are set by the environment and represented by the subject him- or herself*. The most salient and important of these tasks are those that arise through relations to the natural and social worlds. (Ip201)

Following these clues, anti-individualist (or narrow-content) theorists suggest that explanations in terms of the physical syntax of states realising semantic components in entirely intra-individual terms may not be adequate to the explanatory task of psychology. Why then would we have come to be enslaved by individualism?

The only reason for our fascination with individualism and narrow content is the thesis that the states explaining behaviour must be self-contained and relate to proximal stimulus states and that environmental or distal features are not directly formative of the mechanisms designed to represent them and make them accessible to the motor systems of agents (to use Millikan's distinction) which are construed as "producing physicalistic outputs.

Fodor's argument is as follows.

(c1) A creature on twin earth who is my twin and I are duplicates.

(c2) We are therefore physical duplicates of each other.

(c3) Our causal powers are identical.

(c4) There is a causal explanation of the contentful mental states and bodily movements of my twin that does not depend on external factors.

But the current tendencies in cognitive-neuroscience (traceable back to Luria's emphasis on cognitive interaction and environmental effects) tend to make the conclusion unattractive; we find in the brain many circuits designed to lock on to environmental targets and track them. The content of these systems, one could say, is *de re* in that it is not concerned about the proximal content but only about the current status of its distal object. It seems from all that has gone before in the present discussion that the mental states and events of any subject may, therefore, not be able to be individuated in the way required by this narrow-content or internalist theory (in terms of physical or functionalist properties). When this insight is combined with a claim advanced by Kant we are left with a curious twist on Kim's view that proper causal explanations must involve monistic causal chains (1995, Ip125).

Kant claimed (1789, B848) that the practical method of determining the objective content of a representation was by consensus between two observers of the third thing that formed the object of thought (the thing signified). Burge has drawn our attention to the importance of both the natural and social worlds in yielding the objects represented in mental content. We need to add one further ingredient to the picture in order to derive a satisfying view of explanation which combines causality and connectionism in a way recently suggested by Cynthia Macdonald in her discussion of externalism and norms. The further ingredient is the fact that many of the most basic and primitively appreciated elements of our mental world are socially or interpersonally mediated, things like a sense of belonging, discerning the fact that others like me, guilt or rejection, and so on. Other significant mental contents are heavily linguistically mediated like the baroque qualities evident in Handel's music, the porridgy nature of Kant's prose, or the Impressionism that one can detect in a Turner painting.

Macdonald argues for a version of externalism that shares many features with Millikan's biological argument about the proper functions of cognitive processes. She argues that our cognitions and the behaviour they give rise to have to be evaluated in the light of the "fit" between activity and its characteristic intentional objects construed as actual elements in the external environment but then she adds a twist which distances her from

many teleological versions of externalism.

However, the similarity between contentful intentional kinds and functional-biological kinds ends here. Specifically, the attribution of contentful kinds is, whereas the attribution of functional kinds is not, sensitive to both the perspective of others and the perspective of the subject. ... There is no analogue of this dual perspective constraint on functional taxonomy in biology. (1998, p299-300)

In the present discussion of mind and brain I have repeatedly invoked the correction and training by others and the implicit relativisation of one's own contentful acts of perception and thought to which that gives rise. Once again we are forced to conclude that the realm of words or signification created by the socio-cultural reality in which we live and the environmental or external determination of content combine to shape the processing structure of the brain. The causal forces at work here can, at one level, be understood by shifting distributed patterns of information in the brain which fit themselves to the confluence of the conditions of utterance, the parsed utterance, and the intentions of the speaker in a dynamic and relatively holistic way in determining the content of a given experience or mental act. But, at another level, the forces that have moulded those processing patterns can only be discerned by looking at discourse. Therefore if we want to understand psychological explanations we will need to take into our survey these same diverse regularities. That they are not law-like follows from the holistic interaction of two law-like realms (brain and physical world) and a third non-law-like realm (the intentional world of speakers and hearers).

What is more, we must take account both of laws and rules. There are rules governing the use of terms in a language, conventions governing the communication of speakers intentions, and laws operating in the natural (and artifactual) environment (although there is also massive causal particularity in ecologically real situations). At any given moment the network state of the brain may be unique, even though some good-enough types may be identified for the purposes of cognitive neuroscience. Thus strict, physically self-contained, individuating conditions for mental states and events may not be forthcoming. We need not, however slide into pessimism on the basis of such a possibility.

In fact we can find two things from the human activities of the speech community using a natural language:

(i) a way to individuate components of states of affairs that serve as the referents of terms in natural languages in the repeatable pattern markers of semantics or signification; and

(ii) ways of limning the mental connections found in the individuals whose minds are structured according to the semantics of the language or speech community concerned. By using these clues, along with an understanding of the dynamics of interpersonal and cultural influences, we can come to generate an understanding of human psychology that is well-grounded but is not totally specifiable in physicalist terms.

Does this mean that we violate Kim's requirement that explanatory chains be unitary and monistic in order to leave us with respectable epistemic categories. I am sure that it does not leave us quite where Kim would feel comfortable but, in view of Dennett's work and the multiple drafts model of conscious content that may, in any event, have been an unduly optimistic destination to aim for. We can, perhaps, offer a reassuring voice by pointing out that brain processing networks are quite plastic and therefore might be expected to realise within them the artifactual saliences that arise in human forms of life (such as Brooklyn Bridge, classical music, and my family traditions) and that these artefacts should, on any decently Aristotelian theory of mind be represented in configurations of brain activity. Given that none of these things are individual creations the prospects for a self-contained individualistic psychology look grim and given that

they are each heavily cultural in nature, although two are physically realised, the prospects for a self-contained physicalism producing explanations of behaviour totally in terms of physical kinds look equally grim.

Clark has given us good reason to be deeply suspicious of explanations which aim at an output specification produced in the mode suggested by classical material causation accounts of mental explanation.

Such accounts aim to produce a behaviour specification that is structured in the light of internal representations but what we are offered as a result of dynamic cognitive interactionism is not of that type at all. The cognitive processes defined in terms of interactive loops between organism and environment aim at satisfying the multiple constraints I have repeatedly mentioned in an open-ended kind of way so that the human organism pursues a conscious trajectory through the world illuminated by the narrative of consciousness and progressed by active interventions themselves cybernetically coordinated with the ways in which one's environment reacts. Here there is no place for the self-contained internal causal chain leading to the formed output specified narrowly and independently of the context of the activity concerned. It would therefore seem that what Kim and Fodor are after is a target which is not worth aiming at and that we need a more sensitive and informed conception of intentional action than that suggested by a pattern of bodily movements produced as the result of a set of mental states causally antecedent to it.

VI. Resonances with Wittgenstein

Could one say that the decision of a person was not free because it was determined by natural laws?

It is on the face of it extremely curious to them that natural laws were after all general descriptions of what happened, what is going to happen should compel things to happen as they do.(sic)

Therefore to say that the natural law in some way compels the things to go as they do is in some way an absurdity. (Wittgenstein, 1989, 85)

Wittgenstein's lecture on freedom of the will (or at least the notes taken by Yorick Smythies) contain a number of themes that are also evident in philosophical writing on the nature of the will and its properties from the mid twentieth century. He discusses, in addition to the distinction between natural laws or regularities and laws which dictate that we shall do thus and so, the uncertainty about the regularities expressed in psychological laws and the type of determination involved (86), the problems with an unspecified mechanism (88), the specialness of cases where a person's behaviour seems determined by facts beyond their control (91), and the problems about the application to a complex physiological system of certain types of analogy (92) (Hampshire, Thalberg, Vesey).

The type of determination involved in psychological determinism is the first topic to which Wittgenstein turns his attention. He looks at cases where the law involved is a law of nature and concludes that where there is a fairly simple and easily stated law that amounts to no more than saying "it goes as it goes"(p86). In such cases we are often able to predict what something will do as if it were running along rails which determined its activity and then it seems natural to think of the thing as merely following a trajectory laid down in advance and unaffected by anything intrinsic to the object itself. But this is not quite the case with psychology.

Suppose I said 'Our decisions are determined by the circumstances of our education and our whole anatomy. We don't know in what way they are determined. We can't predict except in very rare circumstances and then very roughly. All the same it is

reasonable to think that they follow natural laws and are determined'(p86)

But we tend to say this and slip to the determinist position when the system we are studying shows obvious regularities and gives little evidence of an internal locus of control (to misuse a term from personality psychology). Science, starting as it does with simple causal systems where laws and predictions are accessible and surveyable can mislead us here; "science could only start with obvious regularities, going on on to less obvious regularities"(p87). The problem is that we extrapolate the mechanistic and "soul-less" working of the mechanistic processes we understand, control (externally), and design with what is going on in the more difficult cases where we may have gone beyond a legitimate extension of thinking based on the simple model.

The idea of a mechanism whose working are unclear to us is, in one sense, custom made for the brain which seems a hugely complex set of processes all of which are amenable to simple biological analyses. This leads us to a bullish confidence in our methods and their metaphysics but Wittgenstein registers a voice of caution by reminding us of another direction of extrapolation in which the confidence leads to misplaced conclusions.

You might say (in the case of the thief): 'There is a mechanism here, but a very much more complicated one'.

In the case of electrons one simply gives up. 'No. There are no laws here'.(p89)

Wittgenstein records our mechanistic impulse in the case of psychology as follows "you compare the case with that of a clock – mechanically determined"(p89).

He then develops an interesting analogy by suggesting an engineering situation in which a car is set in motion on a smooth plane surface and because of its internal idiosyncrasies it pursues a "queer path". He notices the difficulties of deciding about responsibility, freedom, and external control in this case. He suggests a further natural case which veers even closer to the idea of currents and patterns of excitation sweeping through the brain and moving us this way and then that.

You sometimes see in a wind a piece of paper blowing about anyhow. Suppose the piece of paper could make the decision; 'Now I want to go this way.' I say 'Queer, this paper always decides where it is to go, and all the time it is the wind that blows it. I know it is the wind that blows it.'

That same force which moves it also in a different way moves its decisions.(p90)

Here we see that if the premise is accepted (and the leaf is deciding) then we have a metaphysical situation very much as envisaged by compatibilists who believe in the liberty of spontaneity. This thought is available to us on the basis of doubts about the kind of causation we might see at work in neural networks given the way that they realise patterns of intentional activity. However Wittgenstein has one or two further thoughts which should be put into the mix.

He asks us to note the specialness of cases of external determination.

We are comparing the case of a human being with those special cases where we would say that a man was determined: where we would say that he thought he was deciding freely, but was actually compelled. Why should anyone be inclined to compare ordinary cases with such a very special case?(p92)

It is clear that we can reliably distinguish cases where an individual is acting in a way that he has determined he will act and for which he ought to bear responsibility from cases in which he is controlled by an external agent. But the idea of psychological determinism, as Wittgenstein suggests, collapses such cases into variants of the same process, some of which are more transparent than others. If we are to resist this it seems we need not only to undermine the tendency to drift towards the deterministic view but also to provide a reason to favour a different metaphysical view.

To do this we need to defuse *the disanalogies that drive determinism*. Wittgenstein

draws our attention to the way in which thinking by analogy can persuade us to take a strange attitude to what we do or what our bodies are engaged in.

He mentions a “kidney advertisement” in which the statement ‘There are 15 miles of kidney to clean’ compares cleaning the kidneys with doing something extremely difficult – which it may not be at all”(p92). In fact, given that we are all doing it continuously without even thinking about it it seems as easy as falling off a log, if not easier. he then sharpens the focus of the point about disanalogy.

‘To understand all is to excuse all’. ‘If you understood the working of his mind and understood all the circumstances as well as you understand a piece of machinery you wouldn’t hold him responsible for this actions.’ I would say: ‘how do you know’? It doesn’t follow any more than that kidneys are difficult to clean.(p92)

Wittgenstein is interested in undermining some metaphysical slides that take us from commonplace observations about human biology to doctrines about key concepts in the moral sciences. We should take from this that a specific type of argument is required to do metaphysics (as much as it can be done at all) and that the findings of science have to be carefully scrutinized before we appreciate their significance for our (metaphysical) view of human nature and the human soul.

All of these arguments might look as if I wanted to argue for the freedom of the will or against it. But I don’t want to.(p93)

Having noted these carefully argued reservations about scientific determinism we can now proceed to evaluate the metaphysics of action and the relevance of Wittgenstein’s remarks on freedom of the will to the extension of the Millian view into a deterministic account of brain and behaviour.

VII. Philosophical distinctions in the problem of freedom of the will.

In discussing freedom of the will, philosophers usually distinguish two different approaches which derive from David Hume. These are the freedom of indifference and the freedom of spontaneity. The Millian account is of the latter kind and argues that lawfully determined causal chains are necessary for a psychologically satisfying account of action linking behaviour to the character of the agent.

It is useful to think of the *Freedom of indifference* as one of two doctrines.

1. *Strong indifference* is the claim that there is no conditioning effect of antecedent conditions on a human action. We might associate such a view with Jean Paul Sartre who claims that you are always free to make up your mind and that no psychological state of yours is at all effective in controlling or causing your behaviour. He argues that your action arises when you confer a meaning on a situation and then act into that situation on the basis of your free choice and what you want to make of it. Motives, the conditions which dispose one to act in a certain way are all predicated on the meaningful choice. “Therefore the caused, far from determining the action, appears only in and through the project of an action”(1958 p448). He concludes freedom appears as an unanalyzable totality; causes, motives and ends , as well as the mode of apprehending causes, motives, and ends, are organized in a unity within the compass of this freedom and must be understood in terms of it.(1958, p452)
2. *Modest indifference* is the claim that whatever antecedents hold and whatever their effect, it is still, in some sense, up to me which way I act and that the antecedent conditions do not causally determine (or necessitate) my action.

Neither of these constitutes (or even clearly indicates the form of) a theory of action,

both are negative claims about the causality appropriate in a discussion of human action, but both negate the idea that a causal nexus physically brings about an action (so both are congenial to Wittgenstein's denial that willing is a type of "immediate non-causal bringing-about" [PI # 613]). However we are left with a need for some way of conceptualizing the fact that my behaviour flows from my character and this seems to lead, inevitably, to the idea of the freedom of spontaneity.

The *Freedom of spontaneity* is a thesis that my behaviour flows from my mental life or is explained by my mental life. It is usually taken to be compatible with Mill's idea that my actions are caused by their mental antecedents and therefore is incompatible with the idea of strong indifference and possibly, depending on one's idea of causality, may be incompatible even with a modest indifference claim.

In order to link the freedom of spontaneity to a determinist theory of behaviour we need to add two further premises to the argument.

C1 *The Causal thesis* states that all genuine explanation is causal explanation (Kim, 1993). I have given rejected this view in the present discussion of connectionism and action explanation.

C2 *The Completeness claim* (or "the doctrine of philosophical necessity" Mill) is that causal explanations form a complete system which, when fully laid out, yields a set of jointly sufficient conditions for all events including human actions.

Mill himself explicitly espouses both of these theses:

given the motives which are present to an individual's mind, and given likewise the character and disposition of the individual, the manner in which he will act might be unerringly inferred: ... if we knew the person thoroughly, and knew all the inducements acting upon him, we could foretell his conduct with as much certainty as we can predict any physical event (1843, p837).

Taken together with the freedom of spontaneity, C1 and C2 give us *Strict Determinism* which is incompatible with any kind of indifference thesis and, most of its adherents would argue, incompatible with a robust conception of freedom of the will of any type.

A Modest indifference thesis which looks as if it might be compatible with some version of the freedom of spontaneity needs the thesis that deterministic causal laws apply most places but do apply to the explanation of human actions. In order to defend this view we would need to defeat either C1 or C2 or to provide a theory of action which embodies a substantial indifference claim.

There are several versions of psychological determinism but, as a stalking horse for the philosophical argument I am about to outline, we need a theory of the general form that Mill set out: the complete explanation of an action (or indeed any mental act) in terms of its causal antecedents can be found in the history of the psychological subject experiencing that state.

This thesis is open to the general worry about deriving causal explanation from historical data. For instance, Wittgenstein objects to Freud's theory as follows: "when we think of causal laws in relation to physical things we think of *experiments*. We have nothing like this in connexion with feelings and motivation"ⁱ

Wittgenstein observes, as would any philosopher of science or metaphysician, that Freud could not in principle defend any causal story because he could not demonstrate the following twin counterfactual claims:

A positive claim connecting cause to effect: *if C (in conditions O) then E;* and

A negative claim about the necessity of the cause: *if not C (conditions O being equal) then not E.*

These are jointly necessary and sufficient to establish the counterfactual claim required for causality: *if C had not happened E would not have happened*.

Even when we relax this requirement to take account of the way we do, in fact, use causal talk and adopt something like Mackie's INUS condition ("an insufficient but necessary part of a condition which is itself unnecessary but sufficient"[Mackie 1975]), we need to have some evidence which supports the counterfactual claim and its lawlike projection rather than some other explanation. This would usually be in terms of an overarching theory of the domain of events of the type in question – behaviour. But here we strike the arguments from connectionism which undermine the idea that there are events of the requisite type to ground the causal explanations.

Like all those who seek to explain actual human behaviour, Freud can claim that, strictly interpreted, Wittgenstein's is an impossible demand because one cannot, in general, satisfy the conditions on the left hand side for events in the past. We cannot switch on and switch off the causal antecedents of historical human events which have led to current psychological states and therefore we must proceed by other means to establish the required counterfactual claim. A determinist might go on to claim that we can infer the truth of the required causal claim because the mechanisms that we have observed make sense of human behaviour and have explanatory value in trying to understand human thoughts, attitudes, and actions. We have therefore got what suffices in many areas of science, an inference to the best explanation. I do not want to get into the relative merits of Freudian and other theories at this point and therefore will assume that the inference to the best explanation is an appropriate characterisation of the kind of justification offered for the truth of such psychologically determinist theories of human action. However, in psychology, we have reason to believe that the effects of the explanation on the critter whose behaviour is being explained must be taken into account (and therefore the role of the subject as interlocutor produces a social version of Heisenberg's uncertainty principle). We cannot therefore just assume that the behaviour of a psychological creature can be explained as one would explain the operations of a non-sentient physical system without some argument to this premise which does not rely on the discredited causal thesis about brain events.

VIII. The causal priority of brain events: evidence and theory.

Some writers would argue that the philosophical niceties are beside the point because we have evidence to hand that there are measurable brain events preceding the mental events which constitute human action (Gillett 1999). Libet, a French physiological psychologist, timed the initial conscious awareness of an intention to move, as reported by the subject, as occurring about 300 - 500 milliseconds later than identifiable brain events reliably correlated with voluntary (responsive) action. The brain event concerned was called a "readiness potential" and I have discussed the methodology in detail elsewhere (1999, 2001b). If Libet were correct, we could argue as follows.

- i. The mental event of intending to act causes the act.
- ii. The physical event precedes the mental event of intending, and is unconscious; therefore
- iii. The physical event precedes and *causes* the mental event of intending.
- iv. Therefore conscious events are determined by their physical counterparts.

I have called this *The argument to physical priority* (APP).

Three philosophical issues, closely relevant to the relation between mental and physical events assumed in the APP, have been raised in the present work.

- a) What is the relationship between conscious experience and human action?
- b) What sense can we make of the concept of simultaneity in play in the argument, i.e.

between two mental experiences and between a mental experience and a physical event?

c) Can we sustain the view expressed in premise (i) of the APP (the mental event of intending causes the act)?

a) I have argued that there are no correlatable events making up the respective mental and physical records but rather that a person uses a set of discursive skills to manufacture a narrative out of the stream of neural activity provoked by the interaction between the embodied subject and the world.

b) This renders otiose the idea of simultaneity between mental and physical events.

c) The discussion of connectionism and causality has undermined the discrete events ontology required for the premises of the argument to physical priority.

The model assumes that an action is a bodily movement causally produced by a mental event called an intention. This latter event is then timed according to a report by the subject and correlated with a brain event preceding the same bodily movement. However, I have argued that it is by no means clear that every action is preceded by some conscious event properly called "an intention to act". There are a number of things I just do in the course of everyday activity and there are some actions which are extended in time – adjusting the brightness of the television screen for instance – which are better thought of as a pattern of activity guided by a conception than a discrete act preceded by a formed proto-act called an intention. In fact when we consider the ways that actions happen, the action theory of Vallacher and Wegner according to which an action is best thought of as behaviour structured by a guiding conception seems much more appropriate. I have outlined this view as showing that a conscious conception provides an overall structuring framework within which a pattern of world-involving behaviour unfolds. That pattern of unfolding activity is an action because it has a narratively coherent place in an adequate understanding of the current intentional relations between the agent and her environment.

In fact McGuigan has gone even further and argued that a neuromuscular model of mind should allow us to conclude that "the human mind is the functioning only of systems of the body" (1997, p368) and should be dealt with in biological and cybernetic terms which do not raise awkward metaphysical questions. Seeing behaviour this way would allow us to work towards understanding the programs that control our behaviour. In service of this aim, McGuigan has carefully documented, over the last 20 years, the fact that "electro-physiologically measured events from the brain, eyes, somatic and speech musculature occur almost simultaneously during a variety of cognitions" (p351). McGuigan concludes from this that covert behaviour is the means of cognition and therefore that thought is a physical activity like any other. Taking into account the fact that all thought is accompanied by covert neuromuscular activity which mimics, "en petit" as it were, what one would normally be doing or saying in relation to the cognitive challenge being tackled, one might more modestly conclude from the same data that cognition is accompanied by and perhaps contributed to by covert behaviour. This would ground a theoretical continuum between overt acting out or speaking out loud and cognition or private mental life in general (in a way that I presaged some years ago [1992, RMT, p113]). In fact such a view is meat and drink to Wittgensteinian conceptions of thought in which thought is a revisiting of techniques and skills used in situations where signs and human activity mutually inform one another. On the view I have recommended cognition is seen as a n activity whereby a structure which records the integration between signs and their use in relation to objects and events in the real world is available to the subject

to give form to a response to a present set of conditions.

McGuigan wants to draw a reductive conclusion and assimilate what we accomplish by representation and cognitive activity to cybernetically controlled physical reactions to what is around us but Wittgenstein hints at the inadequacy of this view when he remarks:

The child, I should like to say, learns to react in such-and-such a way; and in so reacting it doesn't so far know anything. Knowing only begins at a later level.
(OC#538)

Here Wittgenstein is explicitly rejecting a reduction of knowing to acting under cybernetic control from the environment. He credits knowledge, and therefore thinking or cognition with another level. When we explore that thought it emerges that even though thought is a product of actual bodily participation in rule governed human activities in which we interact with the world one has reason to doubt that this is best thought of as a kind of causal process. The doubt arises from the fact that human beings not only are controlled by laws which causally move them but are also subject to quasi-rational or conceptual influences which aim to persuade them to act thus-and-so. Whether or not an individual is persuaded is, as we have seen, a matter which depends on an holistic assessment of that individual's inclinations, projects and lived conscious narrative which "whole" is not, I have argued, able to be assimilated or understood as one might understand a mechanism chugging away according to natural laws.

Helen Longino discusses the problems that arise when "behaviour is treated as a product so that the resulting theory "offers a way of explaining behaviour that does not require appeal to anything like intentions or the self-consciousness of individuals"(1996, p283). She notices that many models (of genetic or biochemical deterministic bent) rely on the idea that there is a serial unidirectional causal flow between input conditions and output conditions and that behaviour is a product of a complex chain that is best likened to a production line in a factory. Given that various intervening processes can be assumed to act uniformly (even though their combined effect may be complex and unpredictable) it seems to follow that an independent variable such as humeral environment in the womb or a genetic initiating process can have a causally deterministic effect on the final product. Longino attacks this general picture of the brain as a serial processing mechanism under unidirectional determining influences which is a necessary part of the picture underlying psychological determinism. She notes the way that the neuronal connectivity of the brain produces a holistic system which exhibits behaviour as a result of its own internal integration and remarks as follows.

To understand the brain as integrating physiological, environmental, memory, and self-inputs, and then to understand decision and action as the result of that integration, rather than as a linear summation of the physiological, environmental and memory vectors, places control of action back in the individual without denting either the biological nature of consciousness or the role of social interactions in the formation of self. (1996, p289)

Longino notices a fact that I have discussed in relation to the causal theory of perception, viz. that the human neuro-cognitive system is not linear but multiply contorted and able to take its own output states as input so as to work complex transformations in which its own activity influences the way in which its activity is produced. Nevertheless one may wish to say that the most that this can give us is a form of the argument for the *freedom of spontaneity* which I outlined at the outset of this chapter and a convinced physicalist determinist might still want to hold the following position.

You say that complex factors within the individual are the holistic cause of behaviour for a human being and that these prominently include social factors but you have not

responded to the basic position espoused by any naturalistic construal of the role of the brain in causing behaviour which is that, in the end, human behaviour is produced by a physical system (however complex) obeying natural laws and that this means that the system is conditioned by its inputs (longitudinal in time and occurrent at any given moment) to produce certain outputs.

Longino is not persuaded by this line of thought and responds (in relation to the theory of brain function she considers most adequate – Neuronal group selection theory) as follows.

The explanatory goal of the theory of neuronal group selection is to answer the question, How must the biological substrates of capacities like the capacities constitutive of moral agency work? This biological theory leaves all the philosophical questions in place. (Longino, 1996, p291)

The philosophical questions that are left in place include the following.

What is the nature of freedom of the will?

Under what conditions might we say that someone acted freely and when might we discern that they did not?

What is the relation between persuasion, coercion, and causation?

Can mental explanation be reduced to physical explanation and therefore regarded as sharing the metaphysical properties of non-intentional events and processes?

Longino has given us further reason to believe the thesis developed in relation to connectionism and mental explanation (Ch 11) that the explanation of action is not straightforwardly assimilated to the mechanistic cause-effect chains that most unambiguously ground determinist metaphysics, but it is not clear that we have seen any reason yet to drive a metaphysical wedge between reasoned action and deterministically caused events sufficient to ground a robust doctrine of freedom of the will.

In the wings of this neurophilosophical debate there also looms the spectre of a more sociological kind of determinism waiting to take the reins from the hands of the biologists should the patten be forced to relinquish them.

IX. Conclusion:

It seems that strict physical determinism of the type most easily derived from mechanistic causal explanations of action is philosophically untenable. What is more the experiments which seem to show that it happens are deeply flawed and show nothing of the kind. We are left with the conclusion that, in a way unique to mental or psychological explanation the socio-politico-cultural context of human beings helps us to understand their thoughts, attitudes, and feelings but that this understanding may not satisfy the conceptual requirements for a deterministic causal explanation and ought to be conceptualised as relating individuals to their life contexts in quite another way. It remains to be seen whether the shackles of physical determinists of a more sophisticated sort can be struck from the ankles of the existential subject who is morally responsible for life choices on the basis of a workable conception of freedom of the will. I will approach that topic through the cognitive neuroscience of rule following and complex social behaviour.

ⁱ Wittgenstein, L. (1966) *Wittgenstein: Lectures and conversations* (ed. C.Barrett) Oxford: Blackwell, p42.