The Retreat from Alienation in Cognitive Science

Paul Loader
PhD Cognitive Science
University of Sussex
April 2012

UNIVERSITY OF SUSSEX

PHD

THE RETREAT FROM ALIENATION IN COGNITIVE SCIENCE

SUMMARY

This thesis examines the relevance of Hegelian-Marxian theory to modern day philosophy of cognitive science. It is suggested that certain key Hegelian-Marxian ideas and themes, such as 'externalization', 'praxis' and 'dialectics', have parallels in modern day cognitive science and that, in some instances a direct connection can be traced from Marxian theory to recent cognitive science, via intermediaries such as Vygotsky, Merleau-Ponty and Levins & Lewontin.

It is also suggested that the overarching trajectory of cognitive science is one that can be usefully understood in Marxian terms as a 'retreat from alienation.' Taking this as one's starting point enables one to unify otherwise disparate perspectives under a single banner. In addition it provides one with a means of evaluating individual accounts, such as Varela, Thompson and Rosch's 'Embodied Mind' and Clark and Chalmers' 'Extended Mind'. Conversely, some recent cognitive scientific accounts, such as Kirsh & Maglio's work on 'epistemic action', offer further illumination of ideas that are ambiguously expressed in Marxian theory.

Statement

I hereby declare that this thesis has not been and will not be submitted in whole or in part to another University for the award of any other degree.

Signature:..............................

Acknowledgements

The Retreat from Alienation in Cognitive Science

Introduction

Justifications

The aim of this thesis is to explore the relevance of certain elements of Hegelian and Marxian theory to modern day philosophy of cognitive science. In particular the suggestion is that contemporary cognitive science is involved in a 'retreat from alienation', where the latter is understood in Hegelian-Marxian terms. Such an exercise might well seem in need of justification both to some of those involved in cognitive science and to some of those with Marxian sympathies. From the point of view of the former it might be questioned why we need to import more "German philosophy" (Brooks, 1991b) into cognitive science given the prominence of Heideggerian concerns in recent years. One answer to this is simply that the parallels and connections between Hegelian-Marxian theory and modern day cognitive science are there to be seen and that it would be difficult to justify ignoring them. However, the aim of this thesis is not just to highlight what is already present but also to suggest that Hegelian-Marxian (henceforth 'Marxian') theory might have a diagnostic role to play. In fact we offer a fourfold justification for bringing Marx and Hegel into cognitive science:

1) There are interesting parallels between Marxian theory and modern day cognitive science.

2) In some cases traceable influence can be found from Marxian theory to modern day cognitive science (via intermediaries such as Vygotsky, Merleau-Ponty and Lewontin.)

3) The trajectory of modern day cognitive science can be understood in a useful way in terms of categories derived from Marxian theory (as a 'retreat from alienation').

4) In some instances Marxian insights can be used to diagnose weaknesses or limitations of certain perspectives in modern day cognitive science (although this should not be taken as implying the inviolability of Marxian perspectives.)

The first two points relate more to factual relations that are held to obtain between Marxian theory and cognitive science, the latter two relate more to a proposed use for Marxian theory as a tool for engagement with cognitive science.

These then are our justifications - to those involved in cognitive science - for bringing Hegel and Marx into the field. As for justifying this project to those with Marxian sympathies - hopefully the thesis itself will go some way towards achieving this. In advance we will merely make the following three points. Firstly if one has an interest in possibilities for social transformation then it is surely a legitimate exercise to try to get a better understanding of the essential nature of the individuals for and by whom such a transformation is to be carried out. As the Marxist biologists Levins & Lewontin argue: "How can we say how society ought to be organized unless we claim to know what human beings are really like?"(Levins & Lewontin, 1985, p.254) It is one of the contentions of this thesis that recent cognitive science shows that what we "are really like" as cognitive agents gels well with Marxian conceptions of the human beings as creatures with the potential to realize themselves collectively through their "own spontaneous activity" (Marx, 1844a, p.80).

Secondly cognitive science as a discipline should be of particular interest to Marxists because of some of the similarities it shares with Marxism. This is so not just because it is centrally concerned with human potential but because its reach similarly stretches over a diverse range of subject areas. Moreover, as with Marxian theory, whilst its interdisciplinary orientation crucially incorporates philosophical concerns, its emphasis on the empirical means that it maintains a healthy scepticism regarding the value of philosophy as a separate activity.

Once this initial shared starting point is acknowledged it then becomes a question of significance whether there is some degree of inevitability in the way that the trajectory of cognitive science is one that has increasingly come to mirror Marxian perspectives. A concrete and all encompassing science of the human being might of necessity be drawn to dialectical concerns and to an endpoint of the unification of the individual as an active, embodied, intersubjective, material being.

Lastly, although we have emphasized the utility of a Marxian perspective in diagnosing some of the limitations of current cognitive scientific accounts, it is also true that recent cognitive science contains much which concretizes what are, in some respects, rather vague ideas in Marxian theory. Thus, for example, certain features of a dialectical perspective find less ambiguous expression in dynamic accounts of cognition, and research into 'epistemic action' gives empirical grounding for a 'praxis' of sorts.

What kind of Marxism?

We have attempted some brief justification to the cognitive scientist, then, for bringing more German philosophy into cognitive science, and to those with Marxian sympathies for why cognitive science should be of particular interest. A further worry that might be had by those who have no axe to grind with German philosophy, or with critical perspectives, is with the *kind* of Marxist outlook that underpins this thesis. 'Marxism' after all is sometimes associated with tyrannical regimes, and dialectics with dogma or doublethink. Meeting this worry involves making reference to some Marxian themes that we have yet to explore (as well as some that will not be explored in this thesis at all). Nevertheless, as there will be little space for developing a political tract in the main body of text, we will here attempt a brief characterization of the variant of Marxism which finds most favour with the author of this thesis, in the hope that the comments made will become more meaningful along the way.

We will begin by emphasizing that the outlook in question is not an 'orthodox Marxist' one. By this we mean first and foremost that it is not to be identified with the rigidified interpretation of Marxism adopted in practice by official communist parties and statist regimes such as those that have existed in the Soviet Union and elsewhere. In saying this we express our opposition both to the mode of organization of such societies and to the ideological expression of this organization. As regards the former our objections are to the imposition of a 'communist' order from above. If a society which facilitates the "free development of all" (Marx & Engels, 1848, p.26) is a real possibility it must, we would argue, be an emergent, non-hierarchical phenomenon, predicated on the "spontaneous activity" (Marx, 1844a, p.80) of the mass of individuals.

As regards the latter - associated with the term 'orthodox Marxism' are certain theoretical perspectives linked together by the general notion that Marxism is 'scientific'. Prominent amongst such perspectives are the set of propositions that make up 'dialectical materialism' (Engels, 1883, 1887 ; Lenin, 1908; Cornforth, 1952) as well as an overarching belief in historical or economic determinism. Our relationship to these perspectives is a little more nuanced. In the case of 'dialectical materialism', as will be seen in the main body of the thesis the position adopted is one that finds much of interest in the application of dialectical principles to material phenomena; and

attempts are made to relate this to later developments in cognitive science, particularly in the theories of dynamics and (to a lesser extent) enactivism. It should be stressed however that a belief in the usefulness of a dialectical perspective as applied to the material world does not imply endorsement of 'dialectical materialism' as a systematic or all-encompassing 'science'. Moreover, as will be seen, some features of a materialist dialectical outlook – for example Engels' interpretation of the 'negation of the negation' - seem fatally ambiguous.

As for 'determinism' – the view of this author is one which sees 'determinism' as a problematic and not necessarily meaningful notion, and which therefore regards any qualified variant of this notion – be it 'historical' or 'economic' – as not usefully endorsable. In the main body of the thesis, insofar as the question of determinism is raised, we ignore such subtleties and simply come down 'against determinism'. This is not to deny, however, that there could be certain principles which determine the range of possibilities in the unfolding of historical events, or that there are definite ways in which our conceptions and actions are inextricably bound up with our material situation.

In place of orthodox Marxism the position taken here has something in common with a number of other Marxian and radical currents. Insofar as emphasis is placed on the notion of 'alienation', and in general on the self-realisation of the subject, the position taken has something in common with the diverse range of perspectives linked with the term 'Marxist Humanism' (e.g. Fromm, 1966 ; Fischer, 1968 ; Dunayevskeya, 1973 ; Levine, 1975). Insofar as the communism envisioned is one that is held to be incompatible with the maintenance of hierarchical power - and insofar as our analysis places emphasis on situated praxis, opposing itself to conceptions of the agent as first and foremost a 'planner' - it has something in common with libertarian communism (Velli,1972 ; Dauve & Martin ; 1974 ; Holloway, 2002) and even anarchism (Bakunin, 1882 ; Malatesta, 1891).

The Structure of the thesis

Having attempted to give some justification for this project and some indication of what type of a Marxian outlook it is allied to, we ought also to give a brief summary of the structure of the thesis, as this is easily lost in the detail of the argument.

In chapter 1 we look at certain key concepts in Hegel's work – emphasizing in particular the notions of alienation, externalization and dialectics.

In chapter 2 we turn to Marx. Our account of Marx is one which places alienation at the centre and continues many of the themes found in our account of Hegel in chapter 1, albeit from a materialist perspective. In addition other concerns are investigated (such as 'praxis' and the division of labour), concerns however which still relate to the Marxian project of surmounting alienation and reconstituting the 'whole man'.

In chapter 3 we offer brief summaries of later Marxist thinkers. Some time is spent looking at Marx's collaborator Engels, and particularly at his detailed elaboration of dialectics. Of the other writers, three (Vygotsky, Merleau-Ponty, Levins & Lewontin) are selected because their work can be seen as forming a bridge between Marx & Engels and modern day cognitive science. The last writer (Sohn-Rethel) is included because of his own Marxian theory of cognition, and because of his novel approach to the idea of 'abstraction' which has some relevance to current debates about 'online' and 'offline' cognition.

In chapter 4 we make the move from Marxian theory to cognitive science. We have said that our aim is to diagnose a 'retreat from alienation' in cognitive science. A little needs to be said therefore about the characteristics of traditional cognitive science and why it might be thought alienated in the first place. This is the purpose of chapter 4, although it should be stressed that the treatment given there is far from exhaustive or systematic.

In chapter 5 we then turn to the 'retreat from alienation' itself. Here we give some indication of the different ways in which modern cognitive science, in its striving for a unified conception of the subject, and in its pursuance of dialectical themes, seems to be heading in a direction that meets such a description.

The last three chapters look in more detail at particular areas of modern day cognitive science in relation to the theme of the 'retreat from alienation'. In each case we find evidence of this retreat, through the incorporation of relevant Hegelian-Marxian themes and perspectives. In each case, however, we also find that the retreat is more or less limited, and a Marxian analysis is brought to bear on this fact. Specifically:

Chapter 6 looks at enactivism. In the first half of the chapter we argue that there are significant parallels between enactivism and autopoiesis and elements of Hegelian-Marxian thought. However close examination of one influential, although idiosyncratic, enactivist text (Varela et. al., 1991) in the second half of the chapter reveals both parallels and divergences. An attempt is made to suggest that, in this instance, the respect in which the account diverges from Marxian accounts – its recommendation of an isolated, passive and contemplative solution to fragmentation - is also the respect in which its argument is at its weakest.

In chapter 7 we turn to Clark and Chalmers' (1998) Extended Mind hypothesis. Here it is suggested that some versions of the extended mind thesis are compatible with, and to some degree are indirectly influenced by, Hegelian Marxian conceptions of 'externalisation'. However it is argued that the 'literal' version of the extended mind thesis presents us with an alienated account of mind. Attention is also focussed on the 'cyborgian' variant of the extended mind thesis which, it is suggested, fails to take account of issues raised by Marxian critiques of mechanized labour.

Finally in chapter 8 we look at Kirsh and Maglio's (1994) analysis of 'pragmatic' and 'epistemic' action. It is argued that epistemic action has much in common with the Marxian notion of 'praxis' – understood as the unity of thought and practice, and indeed goes some way towards providing an empirically demonstrable exemplar of this notion. However it is also argued that the dichotomy between 'pragmatic' and 'epistemic' action is problematic.

Apologies

That then is the structure of the thesis. Having started with justifications we might conclude with some apologies. Firstly although I have given this thesis the title of "the retreat from alienation in cognitive science", this is perhaps misleading insofar as it suggests a comprehensive historical survey of cognitive science from a particular perspective. In fact the range of cognitive scientific concerns explored is fairly restricted and fall mainly within the 'philosophy of cognitive science' domain rather than that of empirical research (although the last chapter broaches one particular area of empirical research.) Secondly, because many of the works referred to in the first three chapters use the gender-specific term 'Man' to specify humanity, and 'men' to describe

individual human beings, the accompanying text, in order to aid readability, also follows this convention in many instances. Attempts have been made to rectify this in certain places but these attempts are probably not thorough enough.

Chapter 1 – Hegel on Alienation, Externalization and Dialectics

Introduction

Our aim in this thesis is to highlight what we see as a 'retreat from alienation' in cognitive science. Before we can do this however we need to look at certain theorists whose work can provide us with the tools and perspectives to diagnose this retreat. In this chapter I will attempt to give a summary of those parts of Hegel relevant, directly or otherwise, to this diagnosis. As with other commentators who have attempted summaries of Hegel (e.g. Weiss, 1974; Taylor, 1975) it is perhaps necessary to plead mitigating circumstances in advance for the likely inadequacy of the resultant text. Firstly because of the notorious difficulty of certain of Hegel's works, and in particular of the *Logic* and the *Phenomenology of Spirit*, much of the foregoing will be interpretative in nature. However hopefully consultation of various secondary sources will mean that the account given is at least not at odds with all interpretations. Secondly, the fact that this account is little more than a snapshot of particular fragments of the Hegelian system might itself be thought indicative of a falsification of Hegel's ideas, at least if we adopt Hegel's own holistic criteria according to which "the true is the whole"(Hegel, 1807, p.11). In defence of this all that can perhaps be said is that, for the purposes of this project, such fragmentation is unavoidable. However insofar as there are also 'degrees of falsification' we might say that it is probably better to have some space devoted exclusively to Hegel rather than make do with a vague impression of him as filtered through the work of Marx.

The account is divided into three sections - 'Alienation in the Phenomenology of Spirit', 'Externalization' and 'Hegelian Dialectics'. This is perhaps something of an arbitrary way of carving up Hegelian theory and much that is relevant to one section may fit just as well in the other. On balance, however, it seems more convenient and aids readability to impose some kind of structure on this summary.

(1) Alienation in the Phenomenology of Spirit

Conceived in its narrowest sense the purpose of this part of our summary of Hegel is to provide an account of alienation as it appears in his writing. The work of Hegel's which is most centrally identified with the concept of alienation (e.g. by Arthur, 1986, following Lukacs, 1975) and the project of transcending (or 'superseding') it – is the *Phenomenology of Spirit* (1807). This is also the work of Hegel's which is generally held to have had the most influence on Marx's account of alienation, although there is some dispute about what passages Marx held to be significant (Arthur, 1983). Our summary of Hegel's concept of alienation will therefore take the *Phenomenology* as its point of departure. It should be noted that, as with Marx, Hegel's concept of alienation is intrinsically linked ('internally related') to a wealth of other philosophical ideas. Some of these ideas will therefore also have to be brought into the picture. Likewise there are parts of Hegel's account which, whilst they do not deal directly with alienation, have had, or are held to have had, an influence on relevant features of Marx's account. These too must be included. Specifically, then, our interest in the *Phenomenology* is limited to a) the overarching themes of alienation and its supersession b) points in significant ways connected with this theme, and c) any other sections or points which are held to have had an influence on relevant features of Marx's own account.

What is the Subject?

Richard Norman describes the *Phenomenology* as "a necessary development through a sequence of forms of consciousness."(Norman, 1976, p.18) Already we are here faced with an obvious question. What exactly do we mean by consciousness? Do we mean an individual human consciousness? Collective human consciousness? Consciousness as some supra human metaphysical entity? i.e. Who or what is the *subject* of the *Phenomenology*? There is perhaps no straightforward answer to this question. The

ultimate subject for Hegel is 'Spirit' ('geist') and the journey outlined in the *Phenomenology* is that of Spirit coming to know itself (Arthur, 1986, p.49). The 'different forms of consciousness' that Norman (and Hegel) refer to can thus be seen as stages on Spirit's own journey towards self-knowledge.

However, in replacing 'consciousness' with 'Spirit' we have in some respects merely moved the problem up one level, for now we are faced with the question of what exactly 'Spirit' is. Much of Hegel's own account seems to indicate that the term 'Spirit' should be taken literally as denoting immaterial being of some kind. In some passages this being seems closely identified with 'mind' (an alternative translation of 'geist') and in this respect we note that Hegel describes himself as having an 'absolute idealist' metaphysics, a metaphysics which he is keen to keep itself distinct from Kant's mere 'subjective idealism' (Hegel, 1817a, p.73). However, if we take him at his word, Hegel was also an orthodox Lutheran (McLellan, 1969, p.2) and since his own time there has been a tradition, amongst some, of interpreting (absolute) spirit religiously (e.g. Taylor, 1975). Either way, whether one takes 'spirit' to mean 'mind' or 'God' (or neither, or something which transcends both) it would seem reasonable to conclude that Hegel's professed belief in Spirit as "self-supporting, absolute, real being" (Hegel, 1807, p.264) is indicative of a belief in the primacy of the immaterial over the material. This, as we shall see in the next chapter, is more or less Marx's interpretation and one that he sought to supplant with a materialist analysis.

However there are various grounds upon which such an interpretation needs to be qualified. For one thing, leaving aside the question of how 'mind' should be understood, Hegel's version of both God and the material world are idiosyncratic. As regards the former, the type of God Hegel thought most conducive to the development of consciousness of freedom in Man was that which was least separate from "human actualities". (Hegel, 1822/31, p.50) Here we already see two issues at play which count against a standard conception of God. Firstly an instrumental concern with the role God plays in the development of human freedom. This ties in with a generally 'historicist' approach to religion (Liebich, 1979), where different objects of worship are linked to different historical social formations, all of which are judged in accordance with how far they facilitate human freedom ("The conception of God, therefore, constitutes the general basis of a peoples character", ibid) Secondly, a conception according to which

the truest understanding of God is that which places him at least remove from Man, the logical conclusion of which would seem to be an approach that sees God as a human creation (a point we shall return to.)

As regards the material world, in contrast to less sophisticated idealists Hegel did not view this as a merely illusory or ultimately unjustifiable construction (Taylor, 1975, p.109) – rather it was a necessary outgrowth of the development of spirit. As one commentator has put it "Spirit must posit itself in objective form" and this objectification is a "positive achievement" (Arthur, 1986, p.51) We might add to this that Hegel was a man of science famously arguing that "to him who looks upon the world rationally, the world in its turn presents a rational aspect" (Hegel, 1822/31, p.11) Scientific endeavour, the discovery of the laws at work in material phenomena, was seen as a necessary step in the progression towards absolute knowledge. Thus, at least in those passages where Hegel is concerning himself with detailed analysis of particular features of the material world (e.g. in *Philosophy of Nature*) the metaphysical backdrop against which such investigations take place could, in the words of a modern philosopher, be viewed as "a difference that makes no difference" (Dennett, 1991. p.125)

If the subject of the Phenomenology is immaterial being, then, this leaves open a range of possible interpretations concerning the nature of this being and of the material world from which it differentiates itself. We ought also to note that some commentators have gone further in offering what might be called – if somewhat problematically - 'non metaphysical' (Lumsden, 2008) interpretations of Hegel. With respect to the term 'spirit' for example it has been suggested that this might denote intersubjectivity (Williams, 1987), "human consciousness" (Stern, 1990, p.43) or the "human community" (Westphal, 2003). Likewise with the material world – Westphal (ibid) has suggested that the *Phenomenology* is grounded in a 'realist epistemology', a position which echoes a declaration made many years earlier by the Marxist Herbert Marcuse that Hegel "means to be dealing with real things." (Marcuse, 1941, p.64).

We will not pursue these matters further here, although it should be noted that ambiguities of interpretation regarding the ultimate subject of the *Phenomenology* and

the world it inhabits will inevitably have some impact on how we are to understand the concept of 'alienation' and its supersession in Hegel's work.

What is Alienation ?

Returning, then, to Norman's 'necessary development through a sequence of forms of consciousness', we might next ask both what the logic underlying this development is and how it is related to the concept of 'alienation'. A clue is found in the concluding paragraph of the Introduction to the *Phenomenology* where Hegel states:

> In pressing forward to its true existence, consciousness will arrive at a point at which it gets rid of its semblance of being burdened with something alien… (Hegel, 1807, p.56)

Here we see one of the central themes of the *Phenomenology* - that of consciousness, in various guises, surmounting the 'otherness' of the world in order it to make itself at home in that world. 'The world' – by which we may mean different things according to the context – appears to 'consciousness' – ditto - as something alien. Consciousness uses various strategies – be they practical or rational – to bring the world back to itself. This theme informs both specific episodes detailed in the book and its overarching narrative.

Already, however, we are simplifying, for consciousness, says Hegel, does not merely find itself in an alien world. Rather it has created that alien world – it *is* that alien world – and the process whereby estrangement from that world is overcome is partly facilitated by consciousness' recognition of that world as part of itself. Hegel outlines this dynamic in the preface to the *Phenomenology*:

> Experience is the name we give to just this movement in which the immediate…becomes alienated from itself and then returns to itself from this alienation, and is only then revealed for the first time in its actuality and truth…The disparity which exists in consciousness between the 'I' and the substance which is its object is the distinction between them, the negative in general….Although this negative appears at first as a disparity between the 'I' and its object, it is just as much the disparity of the substance with itself. Thus what seems to happen outside of it, to be an activity directed against it, is really its own doing, and Substance shows itself to be essentially Subject. When it has shown this completely, Spirit has made its existence identical with its essence... (Hegel, 1807, p.21)

We notice here that Hegel does not feel the need to explain the terms 'alienated' or 'alienation', and the passage is not untypical in this respect. In fact nowhere in Hegel do we find an explicit definition of 'alienation'. This is partly because, as Lukacs has pointed out, the term was common currency at the time of writing. In economics, for example, it was used "to betoken the sale of a commodity" and in natural law "to refer to the loss of an aboriginal freedom to the society which came into being as a result of a social contract" (Lukacs, 1975, p.538). However we should note that Hegel's use of the term extends beyond this, for it implies not simply the 'handing over' of something but both the 'externalization' of something – making objective in the external world that which was previously a mere potentiality in the subject (see next section) – and the consequent subjective experience of estrangement which results from confrontation with this externality as an 'other'. We shall see in the next chapter that both of these aspects of alienation reappear in Marx's own account.

In the above passage, Hegel goes on to describe not only this separation between subject and object but also its reunification. Although the terminology is rather confusing, the suggestion seems to be that there are two stages to this process. Firstly there is mere passive recognition - 'substance shows itself to be essentially subject' which we might loosely interpret as meaning that the object is recognized by the subject as being identical with it, or an extension of, or its own creation; secondly there is active assimilation of subject and object ('Spirit has made its existence identical with its essence'.)

How we interpret Hegel's account of the estrangement and subsequent reunification of subject and object partly depends on what scenario we apply it to and also on what metaphysics we attribute to Hegel. If our object is the whole of the *Phenomenology* and if we attribute a genuinely 'objective idealist' metaphysics to Hegel then this passage can be taken literally as a description of the process by means of which non-material Spirit realises itself. The progress of the *Phenomenology* is that whereby a non material entity posits a world exterior to itself, at first fails to recognise itself in that exteriority, and then through a process of ever more successful encounters is finally reunited with its other. Alienation is superceded and in the process 'Absolute Knowing' has been achieved (Hegel, 1807, pp. 479-493). However if we narrow our focus to specific

episodes in the *Phenomenology* less grandiose – and in some cases less 'metaphysical' - interpretations are possible. Below are three examples of such episodes (not given in chronological order of appearance) which to varying degrees, and in various senses approximate to the notion of a subject's exteriorizing itself, confronting this exteriority as something 'other' and then achieving unification through recognition. We give these examples here primarily as a means of putting flesh on the bones of the idea of 'alienation' and its supersession in Hegel's work, but in some instances they contain additional elements which will become relevant to our account of Marx and/or of cognitive science.

1) The Unhappy Consciousness

Our first example is found in a section entitled 'Stoicism, Scepticism and the Unhappy Consciousness' (Hegel, 1807, pp.119-138). In this section the protagonist is characterised in general terms as 'consciousness' although it is often suggested that the episodes described relate in various ways to specific historical episodes (Norman, 1976, pp.57-58; Stern, 2002, pp. 85-86). Consciousness is shown moving through a number of transitions in its relationship to the world. It begins with 'Stoicism', a position characterised as "withdrawing from the bustle of existence into the simple essentiality of thought" (Hegel, 1807, p.121). Hegel suggests that this form of consciousness, whilst initially liberating, proves to be inadequate insofar as its lack of engagement with the world results in thought that is merely "abstract" (ibid. p.122) and without real content. "Freedom in thought", he argues "has only pure thought as its truth, a truth lacking in the fullness of life" (ibid). We will see later both how this notion of 'abstraction' is important in Hegel's account and (in chapter 2) how Marx took up the idea of a critique of 'pure thought'.

Consciousness then moves on from Stoicism to 'Scepticism'. Here Hegel appears to be envisioning a comprehensive doubting or denial of everything external to the subject, such that: "concrete thinking…annihilates the being of the world in all its manifold determinateness" (ibid. p.123). This is a more active (theoretical) orientation to the world and so is less prone to the insipid abstractions of Stoicism. However inevitably, says Hegel, the Sceptic finds his/herself in a contradictory and untenable position. The

dimensions of this contradiction vary according to what aspect of scepticism one is focussing on. Hegel makes reference to sense-perception, ethics, and more generally to the realm of "contingency" (ibid. p.125). A useful, if simplified, summary of the problematic situation of the sceptic, in terms of a contradiction between his/her denial of the external world and his/her involvement in it, is given by Taylor (1975):

> As embodied subjects we go on living in external reality. We may well declare its non-being, it returns unceasingly and inescapably. So that in fact what we have is an oscillation between a sense of our own self-identity, and an equally acute sense of our dependence on a changing, shifting external reality. (Taylor, 1975, p.159)

It is this uneasy predicament which gives rise to a third mode of being – that which Hegel characterizes as the 'Unhappy Consciousness'. The Unhappy Consciousness is a "divided consciousness" (Hegel, 1807, p.129) which is aware of its own dividedness and "experiences itself as contradictory" (ibid. p.126). In some passages Hegel, in self-referential fashion, appears to identify the Unhappy Consciousness with this awareness itself, rather than the subject of the awareness. Thus, for instance, he asserts that "it *is* the awareness of this self-contradictory nature of itself" (ibid. p.125, my italics). Such a strategy is perhaps to be expected from an absolute idealist standpoint that has no need to clearly differentiate between ideas and those who have them, but we will not pursue this issue here.

The Unhappy Consciousness, then, is self-aware divided consciousness. We will see later that this characterization to some degree anticipates Marx's own depiction of the alienated individual, particularly as 'fragmented' by the mental/manual division of labour (Marx & Engels, 1970b). Hegel then goes on to describe the Unhappy Consciousness' attempts to confront its "deeply divided" (Taylor, 1975, p.160) nature. It yearns for stability, for something eternal and absolute. However whilst it recognizes within itself the potential for such qualities to be realized, paradoxically it does not recognize them as being realizable within itself (Hegel, 1807, p.127). This is because it conceives of itself as, and only as, part of the realm of contingent, changeable nature. Its idea of The 'Unchangeable' (ibid) is therefore projected outside of itself, onto a God who inhabits an "unattainable beyond" (ibid. p.131). This God in turn presents itself to the unhappy consciousness as "an alien Being" (ibid. p.127).

Here we have a characteristic example of the dynamic of alienation as outlined earlier. The subject apprehends certain qualities within itself but does not recognize them as part of itself. Instead it projects them onto an external entity – God – which then confronts it as something alien. As Norman says of this passage, the implication is that "religious consciousness is unsatisfactory just insofar as it fails to recognize in its 'other world' the externalization of its own self."(Norman, 1976, p.60)

Hegel appears to fall short of making the last logical move. The Unhappy Consciousness does not go on to recognize God as its own creation - a move which may not have been explicitly made until later 'Left Hegelian' interpretations of Hegel's work (Mclellan, 1969; Stepelevich, 1983) - and so there is no clear equivalent to 'substance showing itself to be essentially subject'. The rest of the section instead shows the Unhappy Consciousness adopting a variety of strategies for transforming its alienated relationship with The Unchangeable into one "in which it becomes absolutely one with it" (Hegel, 1807, p.130) but none of these strategies is entirely successful. However, as various commentators point out (Taylor, 1975; Norman, 1976; Stern, 2002) the section does finish by alluding to a later transition to 'Reason', where consciousness realises that "in its particular individuality it has Being absolutely in itself, or is all reality" (Hegel, 1807, p.138), a transition which could be interpreted as a 'return to self' from (one form of) alienation.

2) The Individual and Society

A second, and relatively straightforward, account of alienation and its supersession can be derived from passages where Hegel considers the individual in relation to the state. We will look at these only very briefly here. Hegel argues that in societies where the individual is atomised and has no identification with the collectivity or the state, the laws and customs of that society impose on the individual as something other and alien:

> The Spirit whose self is an absolutely discrete unit has its content confronting it as an equally hard unyielding reality, and here the world has the appearance of being something external, the negative of self-consciousness. The world is, however, a spiritual entity, it is in itself the interfusion of being and individuality; this its existence is the work of self consciousness, but it is also an alien reality already present and given, a reality which has a being of its own and in which it does not recognize itself. (Hegel, 1807 p.294)

This is to be contrasted with the situation in societies where the dichotomy between individual and state has been superseded and thus the individual perceives there to be no separation between the two:

> In a free nation, therefore, Reason is truth realized. It is a present living Spirit in which the individual not only finds his essential character, i.e. his universal and particular nature, expressed, and present to him in the form of thinghood, but is himself this essence, and also has realized that essential character. (Hegel, 1807, p.214)

We note in these two passages that equivalents of both 'substance showing itself to be essentially subject' and 'spirit making its existence identical with its essence' are implied. As regards the former, in the first passage the individual is alienated because he does not recognize the institutional paraphernalia that constitute his social surroundings as being of his own making (insofar as they are a human creation.) As regards the latter, in the second passage there is not mere formal recognition of this fact but also a lived experience whereby the (social) environment constitutes an extension of the individual's being. We will return to this latter theme when we look at 'externalization' in section 2.

3) Sense Certainty

A third variant on the theme of alienation and its supersession can be found in passages where Hegel deals with what might be called 'epistemological' issues. In particular in the first section of the *Phenomenology*, entitled 'Consciousness' (Hegel, 1807, pp.58-103), Hegel outlines a transition made by the enquiring mind through different forms of consciousness when confronted with the material world. The first form of consciousness is called 'sense certainty'. This is described as follows:

> The knowledge or knowing which is at the start or is immediately our object cannot be anything else but immediate knowledge itself, a knowledge of the immediate or of what simply is. Our approach to the object must also be *immediate* or *receptive*; we must alter nothing in the object as it presents itself. In *ap*prehending it we must refrain from trying to *com*prehend it. (Hegel, 1807, p.58, author's emphasis)

This type of 'knowledge', which recent commentators (e.g. Westphal, 2003, p.58), have compared with knowledge of sense data, or Russell's 'knowledge by acquaintance' (Russell, 1912, pp.25-32) is subsequently found by Hegel to be wanting. The presuppositions upon which it rests do not hold up. To be immediate and non conceptualized this knowledge must be of particulars only, which Hegel tries to articulate in terms of indexicals such as 'this' and 'here' (Hegel, 1807, pp.58-68) but it soon emerges that knowledge without universals is either very impoverished or not possible. The ambiguity here arises because Hegel does not specify whether 'sense-certainty' is itself a stage in our coming to know the world (e.g. a developmental stage) or a theory about our knowledge of the world which reflection proves to be ill founded, or again an historical phase in the development of Philosophy (Norman, 1976, p.36). Possibly Hegel, in a kind of metaphysical equivalent of 'ontogeny recapitulates phylogeny', might claim that this is a false distinction in the first place. Consciousness, wherever or however it manifests itself, will unreflectively or theoretically take 'sense certainty' as its starting point when trying to get to grips with the external world. This is perhaps another instance where Hegel's refusal to be bound by the usual exclusive categories which underpin analytical philosophy makes summary of his position difficult, a point we shall return to when we look at 'dialectics' in section 3.

Having highlighted the limitations of 'sense certainty', Hegel takes consciousness on through several further stages of orientation to its object. These include 'Perception' which attempts to take Universals seriously but flounders on the problem of whether an object can be understood as a bundle of properties or as a substratum wherein those properties inhere (ibid, pp.67-79), and the 'Understanding' which, in this section, appears to be identified with scientific investigation (Norman, 1976, p.39) and the "realm of laws" (Hegel, 1807, p.91). The Understanding is likewise found to be inadequate, for reasons which we will investigate later, and the journey (in this section) ends with a reappearance of the 'return to self' theme:

> It is manifest that behind the so-called curtain which is supposed to conceal the inner world, there is nothing to be seen unless *we* go behind it ourselves, as much in order that we may see as that there may be something behind there which can be seen. (Hegel, 1807, p.103)

Here then we find some of the same elements that were also present in the Unhappy Consciousness story and in the highlighted remarks on individual and society. Although there is no explicit mention of alienation there is the suggestion of a unity (between subject and world) which is rent asunder – in the rejection of 'immediacy' – such that the world appears as other to the subject. Reunification later occurs when, after some hard work on the subject's side, it discovers itself in the world. Some commentators have interpreted this later transition literally. Marcuse (1968), for example, sees in it Hegel's belief in the revelation of the knowing subject behind the realm of appearances, a revelation which marks the departure of Hegel's analysis from that of Kant which still clings to a belief in a separate 'thing-in-itself':

> Behind the curtain of appearance is not an unknown thing-in-itself but the knowing subject. Self consciousness is the essence of things. (Marcuse, 1968, p.112)

Marcuse, however, being a Marxian oriented materialist interprets this discovery as an Idealist expression of an unacknowledged material impulse to supersede alienation by transforming the world – "Hegel's insistence that the subject itself be recognized behind the appearance of things is an expression of the basic desire of idealism that man transform the estranged world into a world of his own." (ibid. p.110). For Marcuse Hegel is wrong insofar as he literally claims that self-consciousness is the essence of things, but his heart is in the right place insofar as the active appropriation of the object is at the centre of his account.

Other commentators (Norman, 1976; Stern, 2002) have rejected the idea that Hegel need here be taken literally in the first place, whilst likewise stressing that the move to self-consciousness is for Hegel one of 'active' orientation of consciousness towards its object. Norman, for example, suggests that the starting point of 'sense-certainty' is one wherein "knowledge is entirely passive and…consciousness does not work actively on its object." (Norman, 1976, p.30) The journey through Perception and the Understanding to Self-Consciousness is then one of progressive appropriation of the object via the active application of concepts. The sense in which consciousness subsequently discovers itself 'behind the curtain' is thus, on this interpretation, the less contentious one in which "these concepts are the production of consciousness, and therefore in knowing them consciousness knows itself" (ibid. p.45).

(2) Externalization

Under the banner of 'alienation', then, we have looked at various instances wherein a subject is described as estranged or separated from some portion of the world, and wherein overcoming of this separation is contingent upon recognition that the object in question is in some sense oneself or a part of oneself. This latter act of recognition, we might say, implies a separate occurrence whereby some portion of oneself has already been posited in the world, whether or not it is recognised as such i.e. it implies an act of 'externalization'. It is to this notion that we now turn. Before proceeding, however, we should note that the distinction between externalization and alienation is rather problematic. For one thing, as Arthur (1986, p.47) points out, the German word 'Entäusserung' in Hegel is sometimes translated as 'alienation' and sometimes as 'externalization'. In itself this need not present a problem since it is quite common for words in one language to have more than one interpretation in another according to context and use. However we should also note that in this instance certain theoretical issues hinge on the distinction between 'alienation' and 'externalization'. In particular, one Marxian criticism of Hegel is that he fails to make this distinction, so that for him "to produce any real object in the real world is an act of alienation" (Rees, 1998, p.37). We will take up this issue later in this chapter and in the next chapter. For present purposes we will just say that, whatever controversy might attend such a distinction, it remains meaningful to separate externalization from alienation in Hegel's work, at least insofar as the former can be used to describe cases where that which is posited in the world by the subject is not necessarily experienced as something 'other' or 'alien'.

Lord and Bondsman – Externalization through labour

One noteworthy example of externalization can be found in Hegel's account of the 'Lord and Bondsman' (Hegel, 1807, pp.111-119). This account is located in the passage immediately preceding the description of the progression from Stoicism

through Scepticism to the Unhappy Consciousness. As Norman (1976, p.60) emphasizes, this fact is itself significant, for the divided nature of the Unhappy Consciousness is held to replicate the antagonistic relationship between Lord and Bondsman in some respects (Hegel, 1807, p.126). However in the following we will not explore this connection but will treat the account in a more or less free standing fashion. The passage is also particularly significant for the influence it has had on Marxian thought (Taylor, 1975, p.156), and if there is disagreement concerning the amount of importance Marx himself attached to it (Arthur, 1983) it is nevertheless the case that it prefigures and illuminates many of Marx's own concerns. These include the mental/manual division of labour which we shall look at in the next chapter. Finally we should also note that this passage is important because it provides an example of Hegel focussing on *practical* activity. In previous passages we stressed the activity of the subject in coming to recognise itself in the world on Hegel's account. However, the activity alluded to was mainly of a mental sort – various conceptions of the world are grappled with intellectually until unification is achieved. This emphasis on consciousness appropriating the world through the power of thought reflects one of the central assumptions of the *Phenomenology* – that the supersession of alienation is primarily a rational or theoretical task. Making oneself 'at home' in the world is for the most part achieved by getting a better understanding of the world. As Stern notes (Stern, 2002, p.69) this viewpoint is expressed not just in the *Phenomenology* but also elsewhere. In *Philosophy of Nature*, for example, Hegel declares:

> Our intention…is to grasp, to *comprehend* Nature, to make her ours so that she is not something alien and yonder. (Weiss, 1974, p.203, my italics)

Nevertheless, the Lord and Bondsman passage (as well as some passages in other works) show that Hegel was not exclusively concerned with the appropriation of the world through thought but also through practice.

In the section in question Hegel begins by outlining one of the pre-requisites of a genuine 'self-consciousness', which is that it should receive recognition from another consciousness:

> Self-consciousness exists in and for itself when, and by the fact that, it so exists for another; that is, it exists only in being acknowledged. (Hegel, 1807, p.111)

He then spends some time examining the implications of this fact and the complex double edged configurations that must be considered if one is to understand the coming-into-being of mutual self recognition. Very loosely an individual first relates to another individual merely as an object in his world, and is related to in a likewise fashion. However when these individuals reciprocally become aware of their existence as objects for an other, a spiralling process of recognition of self-in-other and other-in-self is set in motion – as Hegel puts it: "they recognize themselves as mutually recognizing one another" (Hegel, 1807, p.112). At this point Hegel introduces an element of antagonism. Although objectively the development of each self-consciousness is dependent on it being recognized as such by another self-consciousness, yet each self-consciousness seeks to preserve its own sense of self-certainty - to "raise their certainty of being for themselves to truth" (ibid. p.114) by annihilating the other as an independent being. Thus a "life and death struggle" (ibid) ensues which is only eventually resolved through an uneven distribution of power. In the resultant set up:

> One is the independent consciousness whose essential nature is to be for itself, the other is the dependent consciousness whose essential nature is simply to live or to be for another. The former is the lord, the other is the bondsman. (Hegel, 1807, p.115)

This set up at first seems to serve the needs of the lord best. The lord ceases to be alienated from the natural world – ceases to confront it as a recalcitrant 'other' – for he has power over it. He uses the bondsman as his instrument – a mere extension of his own being – by means of which he is able enjoy the fruits of the world without having to confront it in its 'immediacy' (ibid). Meanwhile the bondsman seems to be spectacularly disadvantaged for he has no independent existence of his own. The world is to him an object of fear, a fear which has its source in his subjection to the lord. Moreover his very activity is initially a source of alienation for him, "for what the bondsman does is really the action of the lord" (ibid. p.116).

However, the advantages enjoyed by the lord are fleeting. This is so, says Hegel, for a variety of reasons. Firstly, in terms of the 'recognition' that is necessary for the development of genuine self –consciousness, that which the lord derives from the bondsman fails to deliver the goods, for the bondsman has, in the eyes of the lord, been

reduced to a mere object. He is therefore incapable of seeing in the lord that which the lord sees in himself (to paraphrase Findlay, 1977 p.522). Conversely, in a difficult passage, Hegel seems to suggest that the all consuming fear and surrendering of identity to the lord experienced by the bondsman, paradoxically provides the groundwork for the development of a genuine self-consciousness. In a dialectical turn, Hegel argues that the bondsman in surrendering himself entirely to another external power – and so divesting himself of "everything solid and stable" (Hegel, 1807, p.117) in his own particular existence, experiences "a pure universal moment" (ibid) which opens the gateway to the development of genuine self-consciousness. Possibly, as Findlay (1977, p.522) seems to suggest, the idea here is that it is the bondsman, and not the lord, who manages to transcend the bare particularities of his individual existence because he identifies with something beyond himself which he experiences as an absolute power.

Importantly, however, Hegel also suggests that the development of the bondsman's self-consciousness is facilitated through practical activity. We have said that according to Hegel the bondsman does not experience his own activity as belonging to him but Hegel appears to imply that this alienation from activity is a temporary phase. The bondsman, through his working over of the objective world, confronts what he originally experienced as alien and in shaping it achieves power over it, and so becomes at home in the world as an individual in his own right:

> Through work, however, the bondsman becomes conscious of what he truly is…For in fashioning the thing, the bondsman's own negativity, his being-for-self, becomes an object for him only though his setting at nought the existing shape confronting him. But this objective negative moment is none other than the alien being before which it has trembled. Now, however, he destroys this alien negative moment, posits himself as a negative in the permanent order of things, and thereby becomes for himself, someone existing on his own account. (Hegel, 1807, p.118)

In the Lord and Bondsman story, then, we have another variant on the theme of alienation and its supersession. The bondsman confronts the world as an alien and hostile other – partly because he experiences it as an extension of the Lord. However through practical activity he eventually surmounts the otherness of the world and finds himself in it.

It is the manner in which the bondsman does the latter which is of particular interest to us, for the sense in which he finds himself in the world is the sense in which he has *externalized* himself in that world. This applies as much to our previous examples as to this one, but it is clearer in the Lord and Bondsman story because of the emphasis placed on practical activity. It is through action, directed at the world, that the bondsman finds himself in that world. As Hegel puts it: "The bondsman realises that it is precisely in his work wherein he seemed to have only an alienated existence that he acquires a mind of his own" (ibid). The acquiring a mind of one's own is co-occurent, in Hegel's account, with extending one's mind into the world. However it should be noted that Hegel does not seem to be promoting, what in modern parlance might be called, a cognitive developmental thesis. He is not arguing that cognitive powers are developed through manipulation of external nature, although some passages might encourage such an interpretation (Sayers, 2003, p.110). Rather his central concern seems to be with a more fundamental coming into being of self consciousness – of the individual's developing awareness of itself as a being in the world. This 'coming into being' is partly facilitated by an enlargement of the physical territory over which the bondsman has power – that which the bondsman works on becomes an extension of his being despite its externality. As Hegel remarks: "the shape does not become something other than himself through being made external to him" (Hegel, 1807, p.118). Linked with this is a further thesis that that the bondsman creates an image of himself, or "duplicates himself" (Hegel, 1835, cited in Sayers, 2003, p.110) in his product. This idea is not clearly developed in the *Phenomenology* but, as Sayers (2003) points out it receives greater attention elsewhere. One example can be found in Hegel's *Lectures on Fine Art*. The relevant passage is worth citing at some length as we will return to it in the next chapter:

> Man brings himself before himself by practical activity, since he has the impulse, in whatever is directly given to him, in what is present to him externally, to produce himself and therein equally to recognize himself. This aim he achieves by altering external things whereon he impresses the seal of his inner being and in which he now finds again his own characteristics. Man does this in order, as a free subject, to strip the external world of its inflexible foreignness and to enjoy in the shape of things only an external realization of himself. (Hegel, 1835, p.31, cited in Sayers, 2003, p.111)

Similar themes of extension and duplication can be found in earlier writings, pre-dating the *Phenomenology*, which were unpublished until the twentieth century (and so

unknown to Marx – Avineri, 1972). In *System of Ethical Life*, for example, Hegel describes the individual's working over of the object as resulting in a "possession of the product" which entails "the entry of the subsuming subject into the reality of the object" (Hegel, 1802, p.3). Likewise in the *Jena Lectures* Hegel talks of "consciousness making itself into a thing" (Hegel, 1805-6b, p.1). Apropos of our earlier comment concerning the difficulties with the distinction between 'alienation' and 'externalization' in Hegel's work, we might note that in this latter text Hegel does not seem to clearly differentiate between the two, presenting both as a positive good:

> In labouring I make myself immediately into the thing, a form which is Being. At the same time I externalize this existence of mine, making something alien to myself, and preserve myself therein. In the very same thing I see my being recognized. (ibid. p.3)

Whether or not we infer an ultimate equivalence between alienation and externalization from such passages, the main point in all of these accounts seems to be that on Hegel's analysis it is precisely the activity of projecting oneself onto the external world through practical engagement with it that facilitates the coming into being of conscious mind. There is what recent enactivists would call a 'reciprocal codetermination' (Thompson, 2007) here. In Hegel's case the productive individual is reproducing at a micro level the same dynamic that 'spirit' plays out at a macro level, for, as Arthur points out, it is also true that "spirit must posit itself in objective form" (Arthur, 1986, p.51) as part of its own development.

Tools

It should also be noted that Hegel's account of externalization through labour makes some reference to the role of the 'tool' in the process. In early writings Hegel refers to the tool as a "middle term" (Hegel, 1802, p.8) between subject and object. He then rather mysteriously asserts that it "severs objectivity and its own blunting from itself" (ibid). By the latter he seems to mean that the dual sided 'shaping' (of the object and consciousness) that results from the activity of labour, can be conceived as giving rise to not only a physical 'blunting' of parts of the subjects body (e.g. stubbing of fingers) but also a corresponding psychological blunting ("hand and spirit are blunted by it i.e. they themselves assume the nature of negativity and formlessness" ibid.). The

introduction of the tool, however, allows the circumventing of this side effect to some degree.

We will see later how Vygotsky takes up Hegel's notion of the tool as a "middle term" (chapter 3), and also how Marx similarly relates tool use to externalization (chapter 2), in a way that may have been influential on Merleau-Ponty (chapter 3).

Externalization in Property

We have looked, then, at externalization through practical activity (labour) in Hegel's account of the Lord and Bondsman, and in other writings. In some of those passages we noted that externalization involved a 'recognition' of oneself as extended/duplicated in one's product (e.g. Hegel, 1835, p.31). A similar conception of things can be found in Hegel's writings on 'property', which can be viewed as another instance of the externalization of the individual. In this case, however, the social dimension of the 'recognition' involved is more explicitly described. Property is distinguished, in Hegel's early writings, from mere 'possession' because the former implies recognition by another (Hegel, 1805-6a, p.2). Such recognition is not simple apprehension of the fact of possession i.e. it is not just an awareness, on the part of the propertyless individual, that the object in question is now out of one's orbit. Rather it involves projection of certain qualities onto the owner of the object – a consciousness of his significance as a feature of the world:

> In taking possession, the question of recognition comes up again: I take that which could have become his. It merely could become his possession, but it is mine in actuality. His possibility comes after my actuality. He must recognize me as actual. (ibid)

Hegel develops these themes in more detail in later work and particularly in the *Philosophy of Right.* His account there is first and foremost a justification for what is thought to be the "necessity of private property" (Hegel, 1821, p.53), a necessity which, as we will see in the next chapter, is to some degree disputed by Marx.. According to Hegel "mankind has the right to appropriate all that is a thing" (ibid. p.51). However by 'right' here Hegel seems to mean something other than mere contractual obligation. In

fact he has in mind something more akin to a 'need', for he sees property ownership as a pre requisite for personal realisation:

> When I as a free will am in possession of something I get a tangible existence, and in this way first become an actual will (Hegel, 1821, p.51)

Here the talk is of 'will' although elsewhere other mental epithets are used (e.g. "personality must find an embodiment in property"- ibid. p.56). As before the 'tangible existence' that property offers, is as much a tangible existence for others as for oneself.

We find in Hegel's remarks on property, then, another example of the individual's self-realisation through externalization in the material world. This point seems to be implicit in Patten's (1999) suggestion that Hegel's account can be broken down into two separate claims - what he calls the 'self perception' and 'self development' claims:

> The first is the claim that in property I am an object to myself … the second claim is that it is this experience of being an object to myself that allows me to 'become an actual will'. It is through looking at myself in my property that I develop and reinforce the capacities and self-understandings that make up personality. (Patten, 1999, p.147)

However Patten fails to make any connection between 'self-realisation through property ownership' and 'self-realisation through labour', the latter making no appearance in his analysis (although others have made this connection e.g. Marcuse, 1968, p.76). He is thus unable to situate property ownership in relation to practical activity. We saw earlier that the appropriation of the world could either be conceptual (rational comprehension of the world) or physical (labour). The notion of appropriation through property ownership, however, seems to exist midway between the two. Thus:

> Taking possession is partly the simple bodily grasp, partly the forming and partly the marking or designating of the object. (Hegel, 1821, p.59)

The reference to 'marking' here perhaps comes closest to a conceptual appropriation of the world for it appears to denote the merely formal act of designating the object as one's own so that it is recognized as such by others. However we also find references both to the 'bodily grasp' of the object and, most significantly, to 'forming' – which we can assume entails making the object one's own through physical manipulation. Thus to some extent there is a blurring between Hegel's account of property ownership and his

account of labour. Indeed, in certain passages at least, Hegel seems to suggest that taking possession is most conducive to self- realisation the more like it is to labour. Hence, for example:

> The fashioning of a thing is the kind of active possession which is most adequate to the idea because it unites the subjective and the objective. (Hegel, 1821, p.61)

It is difficult to gauge here, of course, whether, in making reference to 'active possession' Hegel conceives of himself as moving beyond the realm of property ownership strictly so called, or as offering an expanded conception of it. Either way, we might say that it at least illustrates the existence of a continuum from 'possession', to 'active possession', to 'active appropriation of the world through labour'.

The Body

Finally, in considering externalization through property ownership we ought also to briefly note how Hegel's elaboration of such appears to reveal a surprisingly objectified account of bodily being. In particular, realisation of the will through property is not, on Hegel's analysis, thought to be qualitatively different to realisation of the will through the body:

> My body is as to its content my universal undifferentiated external existence. It is the real possibility of all definite phases. But also as a person I have my life and body as I have other things, only insofar as they express my will. (Hegel, 1821, p.53)

Thus, in this context anyway, the body is felt to be no different to any of the other furniture of the world – it is a means by which my will can be expressed, but not the only means. Hegel later qualifies this by asserting that "in property my will is not so vividly present as it is in my body." (ibid. p.55) but this very qualification seems to support the idea that there is only a quantitative difference between expression of the will through the body and expression of the will through external media. Both are mere vehicles for the will. We will find a comparable literal mindedness about vehicles for mental contents when we come to look at Clark's 'Extended Mind' in chapter 7.

(3) Hegelian Dialectics

We have so far looked at alienation and externalization in Hegel's work. In this last section we turn to 'dialectics'. For this purpose we shift our attention from the *Phenomenology* to the *Science of Logic*. We should note however that a dialectical approach has been implicit in much of what has already been covered. Many of the transitions we have described or alluded to in the *Phenomenology* and elsewhere – whether they have involved Spirit, the Unhappy Consciousness or the bondsman – have in common the paradoxical idea that consciousness becomes itself by first becoming something other than itself. This can be seen as 'dialectical' in the sense that it implies a state of affairs wherein progress is made through the positing of an opposition, which is then resolved through supersession of that opposition. The antagonism between the bondsman and external nature, for example, is resolved when the former realizes himself in the act of changing the latter. Nevertheless it could be argued that in such transitions what we see is more a dialectic of events than of concepts. Perhaps relevant in this connection is Arthur's (2004) distinction between two types of Hegelian dialectic:

> There are two different types of dialectical theory in Hegel. First is a dialectic of history. Hegel believed there is a logic of development underlying world history. But there is a second sort of dialectical theory found in writings such as the *Science of Logic* and *Philosophy of Right*. This may be termed 'systematic dialectic' and it is concerned with the articulation of categories designed to conceptualise an existent concrete whole. (Arthur, 2004, p.4)

Arthur's distinction between a 'dialectic of history' and a 'systematic dialectic' may not exactly match the distinction we want to make here. For one thing we have not placed much emphasis on questions of world history when looking at the *Phenomenology*. For another, although we likewise wish to shift our attention to the dialectic of the *Science of Logic* (ignoring for now the *Philosophy of Right*), it might be a little misleading to contrast what we find there with what we found in the *Phenomenology* on the grounds that the former is an attempt to "conceptualise an *existent* concrete whole" (my emphasis); for, as we shall see, our account of the *Science of Logic* (henceforth *Logic*) is also to some degree concerned with movement and transition. Nevertheless it is at least true that whereas the dialectical transitions of the *Phenomenology* manifest themselves through the actions of individuals, consciousness, or an anthropomorphized

spirit, those in the main body of the *Logic* are explicitly presented as part of an exercise in categorical analysis.

The Understanding

Turning to the *Logic* then, we might ask what are the main features of the dialectical outlook presented there. In fact it is perhaps easier to begin by noting what Hegel wishes to contrast such an outlook with. Hegel's main concern is to overcome ('supersede') the 'fixity' or 'rigidity' of a conceptual outlook which he identifies with commonsense (and in some passages with certain aspects of the science of his day – Norman, 1976, p.39). This is the conceptual outlook, encountered earlier, which Hegel calls the 'Understanding. Hegel emphasizes the limitations of the Understanding as follows:

> Thought, as Understanding, sticks to fixity of characters and their distinctness from one another: every such limited abstract it treats as having a subsistence and being of its own. (Hegel, 1817a, p.113).

This "fixity" manifests itself most noticeably in rigid adherence to dichotomies. Examples given are many and varied, including 'Being and Nothing' (Hegel, 1817a, pp.124-133), 'Quantity and Quality' (ibid. pp.145-161), 'Identity and Difference' (ibid. p.166-169), 'Cause and Effect' (ibid. pp.215-219) and 'Feeling and Thinking' (Hegel 1817c, p.93). Hegel argues that a belief in the mutually exclusive nature of such categories restricts one to a "partial" or "one-sided" (Hegel, 1817a, p.51) outlook, an outlook whose limited nature means that it cannot provide an adequate account of its object.

We might say provisionally, and in keeping with common usage, that Hegel calls 'dialectical' the style of thinking which is distinct from such one-sided thinking and which therefore has a better chance of grasping its object. Dialectical thought (sometimes also identified with 'Reason' per se) aims to confront the conceptual deliverances of the Understanding so as to undermine their fixity and rigidity:

> The battle of reason is the struggle to break up the rigidity to which the understanding has reduced everything. (Hegel, 1817a, p.53)

38

How, then, does it achieve this? It does so partly by foregrounding the non-exclusivity of dichotomous pairs of predicates. For Hegel, we might say, demonstrating "the finitude of the partial categories of the understanding" (Hegel, 1817a, p.117) involves showing, amongst other things, that apparently contradictory predicates can be said to be true of the object under analysis and that this apparent contradiction is not something to be avoided but is rather to be 'courted' (Acton, 1967, p.23) for it moves thought on to a truer conception of its object.

The Subject/Predicate Distinction

Almost immediately, however, we ought to qualify this, for as it stands it suggests that a) Hegel had no issues with the standard subject/predicate distinction, and b) that Hegel's position could involve jettisoning the 'Law of (Non) Contradiction' and/or the 'Law of Excluded Middle' and therefore jettisoning standard logic. In fact a) is not true (Pippin, 1996; Hanna, 1996) and b) is a matter of some contention (Acton 1967; Pippin, 1996; Hanna, 1996)

To deal with the former point first, the subject/predicate distinction was for Hegel another example of a dichotomy that needed to be confronted. As Hanna argues (1996, p.261) Hegel objected to the ontological presuppositions he thought to be implicit in the formal attribution of predicates to a subject. In particular he objected to the separation posited between that which is predicated and that which it is predicated of:

> One's first impression about the Judgement is the independence of the two extremes, the subject and the predicate. The former we take to be a thing or term per se, and the predicate a general term outside the said subject and somewhere in our heads. (Hegel, 1817a, p. 231, cited in Hanna, 1996, p.262)

The grounds for Hegel's objection to this independence are many and complex - we might briefly just mention three. Firstly – he objects *just on principle* to the assumption made of a separation between thought and its object. As Hanna puts it, for Hegel "thought and its object ...are never ontologically dichotomous" (Hanna, 1996, p.262). Indeed it is just such a predilection for unification, we might say, which underpins the various accounts of the supersession of alienation outlined earlier.

Secondly, and connectedly there is the idea that what is predicated of an object is in reality internal to that object and not just a mere property or quality to be attached to it. In this connection Pippin (1996) alludes to the fact that for Hegel the 'is' of predication and the 'is' of identity are not always conceptually distinct. As is often pointed out (Pippin, 1996; Westphal, 2003) some critics, most notably Russell (1914, p.48-9), have seen this as indicative of a significant flaw in Hegel's reasoning. However Pippin suggests that it is not so much a flaw as a deliberate and conscious redrawing of conceptual parameters on Hegel's part (Pippin, 1996, pp.249-251). That which logicians seek to attribute externally as a property of an object *is* (part of) that object.

Lastly, and also connectedly, there is a holistic dimension to Hegel's objection, as there is to the 'dialectical' outlook in general. Hanna combines this part of Hegel's objection with the two previous objections in a fairly succinct manner when he says that for Hegel the correct conception of a thing "is not something over against the thing but is the thing itself considered in its structured fullness and total relatedness to other things" (Hanna, 1996, p.263). We will return to the question of holism later in this chapter, and also in chapter 3 when we look at the work of Engels and Levins & Lewontin.

Contradiction and the Law of Excluded Middle

As for the question of whether Hegel's position involves jettisoning the Law of (non) Contradiction or the Law of Excluded Middle, and so effectively jettisoning standard bivalent logic, this is a difficult issue. Many commentators have argued that Hegel's conception of 'contradiction' is a non-standard one. Acton, for example, points out that 'contradiction' for Hegel overlaps with other disparate phenomena such as "oppositions, conflicts, tensions and refutations." (Acton, 1967, p.23); and taking this into account we might consider that Hegel is concerned with different matters to that of the formal logician.

Along such lines Hanna (1996) likewise argues that Hegelian logic is not in competition with standard bivalent logic (or 'common logic' as Hegel refers to it.) Rather Hegel's conception of 'contradiction' is one which extends beyond the limited conception of contradiction found in common logic, just as dialectical logic as a whole

extends beyond common logic. This 'extending beyond' involves taking common-logical assumptions as a starting point:

> Hegelian logic is not a competitor of the common logic – not some grandiose alternative logic – but is rather the result of a more adequate ontological reflection on the common logic. (Hanna, 1996, p.281)

Standard logic, with its rigid and static conception of contradiction, is annexed to the restricted outlook of the Understanding. This has its uses in certain domains, particularly formal domains. Problems arise however when we attempt to understand the real world in terms derived from the common logic, as composed of fixed elements that are externally related to each other, for then the implication becomes that "the things referred to by common-logical judgement will be as externally related as the terms in the proposition". (Hanna, 1996, p.278).

One, admittedly limited, way of understanding this might be as follows: In the detached realm of (bivalent) formal logic the mutual exclusivity of a predicate and its negation is a precondition for progress of any sort. It is one of the rules of the game that 'A' and 'Not A' cannot both be true. This is what negation *means* and Hegel, as Hanna points out, is not denying that this is so.

Outside of this realm, however, our predicates succumb to a certain fluidity. In any particular concrete case a sense can always be found where both the predicate and its negation can be applied, with success, to the same object. One (and only one) reason for this is that the meaning of a predicate is always relative to a particular standpoint or perspective. A tree is both 'tall' and 'not tall' relative to other objects in the world, to take a trivial example. Hegel's rejection of 'fixity' and 'rigidity' in the categories with which we comprehend the world can thus, on this (albeit diminished) interpretation, be seen as questioning the assumption that the 'fixed' predicates and exclusive operations of formal logic are always usefully transposable to our interactions with the real world.

We have said, then, that Hegel's approach to contradiction, and to logic as a whole, is not necessarily in competition with standard bivalent logic. However insofar as the latter is felt to be acceptable only when it steers clear of dealings with the real world, we might remark that sense in which Hegel's logic does not seek to supplant traditional

logic is a fairly weak one. This is particularly the case if we consider developments since Hegel's time where syllogistic reasoning has been thought a contender in the race to develop artifacts which engage with and represent the external world.

Mechanism and formal languages

In relation to this latter point it is perhaps significant that Hegel's criticisms of standard logic are sometimes framed in terms of a rejection of mechanism. Thus:

> In judgements and syllogisms the operations are in the main reduced to and founded upon the quantitative aspect of the determinations, consequently everything rests on an external difference, on mere comparison and becomes a completely analytical procedure and mechanical calculation. (*Science of Logic*, p.52 cited in Hanna, 1990, pp.269-270)

Here we see the same critique as before but the reference to "mechanical calculation" adds an additional element, and one which draws our attention to a further feature of a dialectical outlook. We said earlier that part of the problem with common logic was that it failed to take account of the relative nature of predicate application. In expressing matters in this way we were placing emphasis on the cognizing subject. However it is as true to say that, on Hegel's account, there is a similar fluidity, and holism, to be found in the world itself. It is for this reason, as much as any other, that a mechanized logic would be unable to represent the world as it actually is. In this connection Hanna argues that Hegel is not preoccupied with "a Luddite objection to the mere fact of Logical mechanization" but is rather concerned with the illegitimate imposition of the structures of common logic "onto ontological realms where they do not belong, namely the realms of organic relationships, dynamic processes and concrete truth." (ibid. p.270) The world is a dynamic interrelated whole and the logic required to capture it must reflect this fact.

It is on these grounds that Hegel also objects to Leibniz's "language of symbols" (*Science of Logic*, p.684, cited in Hanna, 1990, p.270). Leibniz's "subjection of the syllogism to the calculus of combinations and permutations" (ibid) seeks to represent the world by symbolic means, with a one-to-one correspondence between those symbols and elements of the world. However it is not the case that one can carve the

world up and then reanimate it in this mechanical fashion, for the elements, once "fixed in isolation" (ibid), will not have the same content that they had before.

Holism

We see above that Hegel's critique of traditional logic leads us back to holism. The world cannot be adequately comprehended if its elements are "fixed in isolation". We have already alluded to the holistic nature of Hegel's dialectical outlook in several places. In our introduction, for instance, we noted that for Hegel "the true is the whole" (Hegel, 1807, p.11), and later we saw that Hegel's objection to the isolation of subject and predicate was partly based on the idea that a thing is what it is by virtue of its relationship to other things. We will not spend space examining these ideas further here, particularly as we will be returning to the topic of holism again when we look at Engels and Levins & Lewontin. However we do want to highlight one additional feature of Hegel's holistic outlook - that it does not just apply to static systems but to systems over time. As Rees puts it:

> The whole is not just the end result, the whole is the process of development through which the parts come to constitute the whole. (Rees, 1998, p.46)

Thus an adequate conception of the object is not one that is based on a 'snapshot' of it at a particular moment, no matter how detailed and inclusive that snapshot might be. Holism has a temporal dimension and understanding one component of a whole does not just entail relating it to other co-existent elements, but also entails relating it to earlier and later (potential) events.

Unity and Sublation

We have said that Hegel rejected the fixity and rigidity of 'The Understanding' and that this rejection was partly made for dynamic/holistic reasons and partly because of an antipathy to exclusively dichotomous predicates. Returning to the latter point we should note that there is more to Hegel's analysis of oppositional predicates than the observation that a pair of such predicates might both be truly applied to an object (or

process). There is also the further suggestion that in applying them in this way we (at least in some instances) make explicit their co-dependence or 'unity':

> The speculative stage, or stage of Positive Reason, apprehends the unity of terms (propositions) in their opposition. (Hegel, 1817a, p.119)

To an extent the highlighting of this 'unity' can be seen as an exercise in conceptual clarification not that dissimilar to the project of ordinary language philosophers who wished to discharge certain philosophical problems by showing how they were rooted in a misunderstanding of everyday concepts (a point also made by Norman, 1976 and Stern, 2002). For example the insight that 'presence implies absence' can be used to lessen the force of particular philosophical puzzles which still resurface in contemporary philosophy of cognitive science. A case in point is Noë's problem of 'the whole cat':

> A cat sits motionless on the far side of a picket fence. You have a sense of the presence of a cat even though, strictly speaking, you only see those parts of the cat that show through the fence. How is it possible that we can in this way enjoy a perceptual experience as of the whole cat?   (Noë, 2004, p.60)

This puzzle has most force if one imagines that the notion of a 'whole cat' – a cat that, perceptually speaking, is entirely and only present - is meaningful in the first place. From a Hegelian perspective, however, such a notion is something of an abstraction; it implies pure 'presence' without any 'absence' and so misses the fact that seeing a cat just is the experience of having portions of a cat 'present' to us whilst other portions are 'absent'. From the point of view of perception, then, a 'whole cat' is necessarily a 'unity' of presence and absence.

In some passages such unities are linked with the notion of "supersession" or "sublation". Hegel employs the German equivalent of the term 'superseding'('Aufheben') to give a sense of the dual nature of a movement of thought. 'Aufheben' implies both 'negating' and 'preserving' (Hegel, 1807, p.68). That which was false in the original conception of things is negated whilst the true element is preserved and carried forth into a higher unity. Thus, with the case of contradictory categories, that which posited mutual exclusivity on either side of the divide is

jettisoned, and the resultant synthesis[1] gives us a more accurate representation of a state of affairs. In some instances – though not all – this synthesis involves the positing of a third term which is held to incorporate the positive aspects of both of the previous terms.

Below we will give three examples of such progressions from the *Logic*. It is important to bear in mind that Hegel considered the transition from one concept to the other to be as significant as the concepts themselves. This was another reason for his objection to standard bivalent logic which he thought held "two determinations over against one another and has in mind only them but not their transition, which is the essential point and which contains the contradiction" (*Science of Logic*, p.441, cited in Hanna, 1996, p.278). As always there is some ambiguity whether the 'transition' to which Hegel refers is a transition only in thought or in the objective world also, or both (and as usual we have to add the rider that this distinction is itself problematic with Hegel). Nevertheless we can say that it refers at least to the former. Hegel is concerned in the *Logic* with the transitions thought makes when presented with a particular concept. Thus Hegel's criticism of the formal contradictions of bivalent logic hinges partly on the notion that movements in thought – rather than the results of thought – are relevant concerns for logic. It is for such manoeuvres, as Hartnack points out, that Hegel was accused of "bringing movement into logic" (Hartnack, 1998, p.14).

Being, Nothing and Becoming

Our first example of such a transition is that outlined in the 'First Subdivision of the Logic' (Hegel, 1817a, pp.123-169) as a progression from 'being' to 'nothing' and then to unification in 'becoming'. Some of the obscurity of this movement might be removed if it is explained beforehand that Hegel's intention was to concentrate on "the world of simple essentialities freed from all sensuous concreteness" (Harnack, 1998, p.15). Hence he wished to start out with talk of 'pure being' freed from any concrete determination.

---

[1] We should, however, be wary of attributing to Hegel an explicit adherence to a triad consisting of 'thesis, antithesis and synthesis'. Such a construction rarely if ever makes an appearance in his work (Mueller, 1996), although some commentators (e.g. Rees, 1998) see it as implied.

In considering 'being' per se, says Hegel, one is inevitably led on to consider its opposite 'nothing', for 'being' per se denotes nothing (Hegel, 1817a, p.127). In considering 'nothing' however we are drawn into considering its identity with being, insofar as both lack a referent. At this point Hegel takes a step back and notes that this movement back and forth (between 'being' and 'nothing') is itself significant and constitutes a third category which unites them both ('becoming' – being becomes nothing and vice-versa). 'Becoming' is thus the unity of 'Being' and 'Nothing' (ibid. p.128)

Whatever we might think of this reasoning, it has additional significance for us insofar as it highlights the importance Hegel laid on the category of 'becoming'. We have already alluded to this to some extent when we noted that Hegel's holism had a temporal aspect. Here Hegel reinforces this idea with reference to the pre-socratic philosopher Heraclitus:

> Becoming is the first adequate vehicle of truth….When Heraclitus says 'All is flowing' he enunciates Becoming as the fundamental feature of all existence. (Hegel, 1817a, p.132)

Hegel, then, understood the world in terms of process and the realization of potential. It is an intrinsically dynamic outlook, a point which is well made by Rees when he suggests that the supplanting of 'being' and 'nothing' by 'becoming' involves "replacing two static concepts with one dynamic concept" (Rees, 1998, p.50).

Quantity, Quality and Measure

A second example of the unification of two concepts via the interpolation of a third concept can be found in passages where Hegel deals with 'quantity', 'quality' and what he calls 'measure'(Hegel 1817a, pp.135-161). Hegel goes into considerable mathematical detail in these passages, particularly when addressing the concept of 'quantity' in relation to other connected concepts such as 'unity and sum', 'ration', magnitude', 'quantum' and 'degree'. We will bypass this detail and merely note in a general way that for Hegel the concepts of 'quantity' and 'quality' are reciprocally related and this relation is expressed through use of the term 'measure'. In fact, as Inwood (1989) also notes, this term is itself used in a variety of ways and perhaps with

some equivocation (e.g. "God … is the measure of all things" – Hegel 1817a, p.157); but the key idea seems to be that the 'measure' of a particular quality is that quantity of it beyond which it passes into being a different quality:

> This process of measure, which appears alternately as a mere change of quantity and then as a sudden revulsion of quantity into quality, may be envisaged under the figure of a nodal (knotted) line. Such lines we find in Nature under a variety of forms. (Hegel, 1817a, p.160)

Whereas elsewhere the question of whether dialectical transitions are a property of thought only or of the world also, seems bound up with metaphysical issues concerning the relation between the two in Hegel's overarching system, here Hegel states unambiguously that this dialectical relation applies to natural phenomena. Examples given include "increase or diminution of the temperature of the liquid water" (ibid. p.159) resulting in ice or steam, as well as standard examples from Greek sorites paradoxes ("whether a single grain makes a heap of wheat or whether it makes a bald-tail to tear out a single hair from the horses tail"- ibid. p.159).

Cause, Effect and Reciprocity

A third example of the sublation of two concepts by a third is that of 'cause', 'effect' and 'reciprocity'. Hegel puts forward a variety of arguments for the inseparability of 'cause' and 'effect' (Hegel, 1817a, pp.215-217), These include a) that one cannot consider 'cause' without considering 'effect' b) that conceptually an effect is the cause of a cause's being a cause, and vice versa c) that in certain specific cases the terms 'cause' and 'effect' have the same denotation – for example: "the rain (the cause) and the wet (the effect) are the self-same existing water" (ibid. p.216), and d) that in some cases certain features of the object affected facilitate the resulting effect and so can likewise be seen as a cause. From such considerations, and particularly from d), Hegel concludes that "in this manner causality passes into the relation of Action and Reaction, or Reciprocity" (ibid. p.217). Hegel sees 'reciprocity' or 'reciprocal action' as providing a more adequate description of causal relations per se - "Reciprocal action realizes the causal relation in its complete development" (ibid. p.218). As was the case with Quantity, Quality and Measure Hegel proceeds to give real world examples, although in this instance they do not illustrate the transition itself but the appropriateness of the resultant synthesis. Thus for example only an explanation in

terms of reciprocity can answer the question of "whether the character and manners of a nation are the cause of its constitution and its laws or if they are not rather the effect." (ibid. p.219) and likewise with living organisms: "the several organs and functions similarly seem to stand to each other in the relation of reciprocity" (ibid.)

It should be noted that the notion of reciprocal interaction, although first developed in detail in response to specific puzzles about causation, is in fact central to Hegel's whole system. Burbidge (2006) points out in this respect that 'reciprocity' is implicit in many of the conceptual progressions in the Logic, including those we have addressed:

> Being and nothing reciprocally interact in the concept of becoming, as do…quantity and quality in the concept of measure. (Burbidge, 2006, p.79)

Thus, 'reciprocal action' is both a move in a particular dialectical progression and a central feature of the dialectical approach itself.

The dialectic within dialectics

We have given, then, three illustrations of what we might call 'dialectical progressions' from the *Logic*. We should note here that, whilst in certain passages the term 'dialectic' is used in a general way to denote the entire modus operandi of developing thought (e.g. Hegel, 1821, p.37), when Hegel is being more precise about the mechanics of these progressions he reserves the term for only one of their stages. Hence in one passage the three 'moments' in a movement of thought are given as "a) the Abstract side, or that of the understanding; b) the Dialectical, or that of negative reason; c) the Speculative, or that of positive reason" (Hegel, 1817a, p.113) Burbidge gives a fairly succinct explanation of Hegel's meaning here:

> Understanding isolates a concept as a unit of thought. It integrates a synthesis into a unity and marks out its determinate limits. Dialectical reason is the transition of thought that moves from a concept to its contrary. This direct shift, suggests Hegel, is the inevitable result of isolating a term from its context through understanding. Speculative reason is the work of reflection: it looks back over what has happened, sees the various moments as well as their connections and identifies the significance of what has gone on. (Burbidge, 2006, p.41)

Given that we have already characterized the Understanding as all that is inadequate and one-sided in thought it is perhaps confusing that it is depicted here as 'integrating a synthesis into a unity.' The rationale behind this seems to be as follows: The initial one sided deliverances of the Understanding are countered by the intervention of the negative at the dialectical stage (e.g. it is shown that 'cause' also implies 'effect'). The speculative stage is then concerned with synthesis, giving an account of the co-dependent nature of these two components (in this case by positing 'reciprocity' as unity of cause and effect). The cycle then returns to the Understanding which once again has a single, discrete term as its starting point ('reciprocity'), albeit one which now incorporates elements of the previously two separate terms.

It will not be necessary to keep in mind this narrower conception of the term 'dialectic' for the remainder of this thesis. However it is perhaps worth noting the similarities between the above account and that of the 'negation of the negation', which we will encounter when we look at Engels' work in chapter 3.

Abstraction

Some mention ought also to be made here of Hegel's use of the term 'abstract' which we have already encountered in a few places. In an earlier passage we saw how the "pure thought" of the Stoic was regarded as 'abstract' and so lacking in real content. Likewise we have more recently seen how the limited outlook of the 'Understanding' is also considered abstract. As might be expected, clarification of this term is not straightforward. This is partly because Hegel does not always use this term in 'the same' way. Grier (1990), for instance, notes six possible interpretations of the term. Matters are further complicated (as Grier also notes) by the fact that the fixed points against which one might hope to define a concept such as 'abstract' cease to be fixed in Hegel's account. Thus for example, one might think that the distinction between 'abstract' and its opposite 'concrete' could be anchored in some way to the traditional philosophical distinction between 'universal' and 'particular', but this particular distinction is itself developed in an idiosyncratic way in Hegel's account, with both concepts themselves declared to be 'abstractions' in at least one passage (Hegel, 1821, p.18).

Nevertheless, simplifying somewhat we might say that sense of the term we are particularly interested in for current purposes is the fairly pejorative sense where the word is used not merely to describe a certain detachment from reality but also to imply that the deliverances of the understanding do not - to adopt a common metaphor – 'carve reality at the joints.' The concepts of the understanding are 'abstract' insofar as their very fixity means that they are unable to fix the phenomena at which they are directed. As Norman puts it an idea is abstract "not in being over generalized, but in being cut off – abstracted – from its relations to the other concepts which make it intelligible." (Norman, 1976, p.43) and later he adds that abstractions are "concepts which are incomplete in themselves and require to be taken as moments in a totality" (ibid. p.64). The concept of 'abstraction' then, is also one which can only be understood in relation to Hegel's holistic outlook.

The Object of Dialectics

At the beginning of this chapter we speculated about the subject of the *Phenomenology*. As we approach its end we might briefly address the issue of the object of the *Logic*. To what does 'dialectics' apply? Novack suggests there are three possible answers to this question. The first is that dialectics applies to nothing because it is "sheer metaphysics" and "meaningless verbiage" (Novack, 2002, p.189). The second is that it applies to "mental or social processes but not nature" (ibid). The last is that it applies to everything including nature. Which of these is true of Hegel?

Burbidge's account seems to imply that the last of these positions must follow from Hegel's metaphysics. Taking as his starting point Hegel's declared intention of superseding the Kantian dichotomy between the world as we experience it and the world as it is in-itself, he suggests that on Hegel's analysis therefore "the study of the way pure thought thinks can also be read as an exploration of the ultimate principles of reality." (Burbidge, 2006, p.15) and later continues "the concepts pure thought thinks, as well as the transitions it makes, capture the structures and processes of reality" (ibid. p.35). However Burbidge has perhaps blurred together two separate possible positions here. All that is needed to supersede the Kantian dichotomy is a position which states that dialectical reasoning helps us to understand the world as it is in-itself. This is not the same as the position which states that the dialectical reasoning helps us understand

the world as it is in-itself because that world is objectively dialectical – i.e. that it itself contains dialectical transitions etc.

In fact, although there *is* a fundamental ambiguity concerning the over all ontological status of the transitions presented in the *Logic*, we do not need to resort to metaphysics in order to establish Hegel's position on this question. We have already said that for Hegel the world was to be understood in a dynamic and holistic fashion, and we have seen, for example, how he thought that the transition from quantitative change to qualitative change revealed itself in natural phenomena. Thus there is little doubt that he thought that dialectical processes were 'out there' in as strong a sense as is possible, given an absolute idealist starting point. This realism about dialectical processes is evidenced by a smattering of other comments throughout the *Logic* (particularly in footnotes) dealing with empirical matters. We note for example his general statement that the purpose of 'Dialectic' is to "to study things in their own being and movement" (Hegel, 1817a, p.117). We note too his observations that "Everything around us may be viewed as an instance of dialectic" (ibid, p.118), "Life as life involves the germ of death" (ibid. p.117), "Matter…is the unity of attraction and repulsion is gravity" (ibid. p.204) "General experience shows us the extreme of one state suddenly shifting into its opposite" (ibid. p.118). As we shall see in chapter 3, later writers such as Engels (1883, 1887) and Levins & Lewontin (1985) have taken such statements as a cue for the development of a full-fledged natural dialectics.

Conclusion – Dialectics and Alienation.

In this chapter, then, we have looked briefly at the concept of 'alienation' in Hegel's work, conceived most generally as the separation of the subject from the object and the consequent presentation of that object as something alien to the subject. Alienation is 'superceded' when the subject recognizes itself within that object and thereby achieves a unity with it. We have also looked at the concept of 'externalization' which, on our analysis at least, is a process implicit in but not identical with alienation. A subject externalizes itself by extending and/or duplicating itself in its environment. In doing so it simultaneously makes its own being more substantial. Lastly, we have looked at

Hegel's dialectics. We have said this is characterized by a number of features including (but not exhausted by) holism, an anti-dichotomous approach, a dynamic view of the world and a belief in reciprocal action.

We might conclude by noting that, on Hegel's analysis, dialectical thought and the supersession of alienation go hand in hand. Dialectics, because of its inherently unfragmented and anti-dichotomous nature, is the style of thinking most suited to the unification of the subject with the object. This point is well made by Stern who argues that in the *Phenomenology*:

> …consciousness is faced by apparently intractable difficulties in making the world a 'home', until at last it comes to recognize that what underlies these difficulties is its failure to think dialectically (Stern, 2002, p.23).

In later chapters our focus on the 'retreat from alienation' in cognitive science will take as given this co-dependence between dialectical styles of thinking and the project of developing a unified 'unalienated' conception of the agent. Before considering such matters, however, we must turn to Marx.

Chapter 2 – Marx on Alienation and Praxis

Introduction

In chapter 1 we looked at Hegel, focusing not only on material that deals explicitly with alienation but also on related material which we grouped together under the headings of 'externalization' and 'dialectics'. We now turn to Marx. Here again our analysis will extend beyond those elements of his work which might be thought to deal specifically with alienation. Our rationale for this is essentially the same as with our treatment of Hegel. The concept of alienation is internally related to a wealth of other concepts – which concepts are also of relevance to our analysis of cognitive science. Nevertheless the interpretation of Marx put forward is still one that views alienation as central to his account. In his philosophic work at least, all roads lead to alienation and its supersession.

It should perhaps be made clear that such an interpretation has not just been manufactured for the convenience of this thesis but rather represents a particular strand in Marxian thought. Whereas some commentators (e.g. Althusser, 1969) have seen Marx's interest in 'alienation' as being confined to a youthful 'Hegelian' period which was not representative of the main corpus of his work, others have seen it as central to his outlook. (e.g. Fromm, 1966; Fischer, 1968 ; Mészáros, 1970). In favour of the latter interpretation it is often pointed out (e.g. Mészáros, 1970 ; Humphries, 1971; Mclellan, 1979), that references to 'alienation' are not restricted to Marx's earlier works. We will not debate this issue further here however.

Turning Hegel 'right side up'

Many of the themes we found in Hegel's work, including the major ones of alienation, externalization and dialectics, reappear in the works of Marx. This is not surprising as much of Marx's philosophy was developed in response to Hegel. He was not alone in this but was following a path travelled by his 'young Hegelian' contemporaries whose work also had an influence on Marx's in various ways (McLellan, 1969; Stepelevich,

1987). Our concern here, however, is not with this latter relationship but with the relation pertaining directly between Hegel and Marx.

A useful summary of this relationship, and one which highlights ways in which Marx is in simultaneous agreement and disagreement with Hegel, can be found in the hedged endorsement he gives of the stratagem adopted in the *Phenomenology*:

> The mystification which dialectic suffers in Hegel's hands, by no means prevents him from being the first to present its general form of working in a comprehensive and conscious manner. With him it is standing on its head. It must be turned right side up again, if you would discover the rational kernel within the mystical shell. (Marx, 1873, p.15)

Here Marx has taken Hegel at his own word and interpreted the subject of the *Phenomenology* as being that of 'Spirit' or 'Mind'. This is the 'mystification' to which he refers. Hegel, according to Marx, has made the mistake of taking one aspect of human existence and placing it on centre stage as an entity in its own right – he has taken "the life process of the human brain i.e. the process of thinking" and has transformed it "into an independent subject" (ibid, p.14.)

Nevertheless, says Marx, if one disregards the purported subject of the *Phenomenology* one can see the essential truth in Hegel's position. Hegel is right in his diagnosis of an alienated subject. As he says elsewhere, Hegel "grasps the alienation of man, even though man appears only in the form of mind" (Marx, 1844a, p.100). Moreover, he is broadly correct in his description of the overarching process whereby this subject realizes itself and surmounts alienation through the resolution of antagonisms presented by the world. Turning Hegel 'right side up' however means restoring the subject to a fully fledged human being – "an embodied, living, real sentient objective being" (ibid, p.104) - and recognizing that the self-realization of this embodied subject is a practical, rather than rational, task.

The Self-Creation of the Embodied Subject

In fact, as we saw in the previous chapter, there are passages where Hegel himself also presents self-realization - or "self-creation" (ibid, p.101) - as a practical rather than

merely rational matter, as for example in the story of the Lord and Bondsman or in his analysis of labour in his early unpublished writings. Marx would not have been aware of the latter and there is some debate concerning his familiarity with the former (Arthur 1983 ; Sayers, 2003). Nevertheless, even if these are acknowledged, the point for Marx would seem to be that, in the *Phenomenology* at least, such concrete instances are presented as mere moments in the self-development of an abstract 'self-consciousness' which ultimately realizes itself in an ideal, disembodied fashion (Marx, 1844a, pp.100-101). Marx wishes to supplant this idealist account with a materialist one.

What, then, are the constituent elements of the practical self-creation of the embodied subject in Marx's account? In answering this we should begin by noting that, for Marx, man is an essentially productive creature. He is, as Avineri (1968, pp.65-95) puts it, 'Homo Faber'. Implicit in such a conception is not only the idea that man's physical survival is dependent on labour, or even just that he finds satisfaction in labour, but that it is through this activity that man becomes what he is – he is the "result of his own labour" (Marx, 1844a, p.101). We should note that 'labour' here is not to be considered merely in terms of the narrowly prescribed activities that are allotted to the individual in the form of 'work' in modern industrial societies. Rather by 'labour' is meant a more generally conceived working over of the natural world for diverse ends - a working over which would ideally manifest itself as "free conscious activity" (Marx, 1844a, p.82 & see later). This notion of the 'self-creation through labour' incorporates several theses which, albeit in a somewhat artificial way, can be considered separately.

Firstly there is what we might call a 'developmental' thesis – the idea that the agent ('man') develops by acting on the material world:

> Through this movement he acts upon external nature and changes it, and in this way he simultaneously changes his own nature. (Marx, 1867, p.283)

There is, then, a reciprocal interaction between Man and the material world – Man effects changes in the material world and in turn his 'nature' is affected by these very changes (and vice versa). There is a "dynamic relationship between man and nature" whereby "both poles are transformed" (Arthur, 1986, p.11). We should note that this transformation has a cognitive dimension, and thus one implication is that the cognitive

faculties of the agent are developed through the alteration of the external world in production of an artifact. As Fischer puts it - "As men changed the world they expanded and refined their ability to know it, and the growing capacity for cognition again enhanced their ability to change it " (Fischer, 1968, p.p.152-153). This idea has been developed in various ways by other Marxists, including Marx's contemporary Engels (1883, 1887), and later by the Russian Psychologist Vygotsky (1930, 1978). We shall return to these two theorists in the next chapter.

Secondly there is implicit in the notion of 'self-creation through labour' what might be called a 'duplication' thesis – the agent's sense of who s/he is, is reinforced by seeing him/herself in the works s/he has created:

> It is in the working over of the objective world that man first affirms himself as a species-being. This production is his active species-life. Through it nature appears as his work and his reality. The object of work is therefore the objectification of the species-life of man; for he duplicates himself not only intellectually, in his mind, but also actively in reality and thus can look at his image in a world he has created. (Marx, 1844a, p.82)

Here, as Sayers (2003) also point out, we see clear echoes of Hegel's idea, outlined in the previous chapter, that "man brings himself before himself by practical activity". For Marx (or at least the young Marx), however, this process was linked with the idea of 'species being' (and 'species life'). This perhaps confusing, and variously used, term was originally derived from Feuerbach who was one of many to ask the question: "What constitutes the essential difference between man and the animal?" (Feuerbach, 1841, p.97). According to Feuerbach the answer lay in the idea that only human beings are able to make their own species into an object of thought. This, on his analysis, provided the precondition for other qualities more typically taken as key indicators of humanity, such as consciousness and empathy:

> Strictly speaking, consciousness is given only in the case of a being to whom his species, his mode of being, is an object of thought… Man is in himself both "I" and "You"; he can put himself in the place of another precisely because his species, his essential mode of being – not only his individuality- is an object of thought to him. (Feuerbach, 1841, pp.97-98)

In Marx's case we might interpret 'species being' as denoting Man's essence as a *socially* productive animal. This may be something of a simplification but to go into

any greater detail at this juncture would take us on too much of a detour. The point for present purposes is that with Marx, as with Hegel in the Lord and Bondsman parable, there is an intersubjective dimension to the self-creation of the subject - "the relationship of man to himself first becomes objective to him through his relationship to other men." (Marx, 1844a, p.84).This intersubjective dimension expresses itself not passively but actively, for it is through productive activity that the relationship of agents to each other and themselves is established and developed.

As regards the 'duplication' of the agent in his product, we might suggest two different 'social' interpretations. Firstly we can say that the individual duplicates himself for others and that his sense of identity is therefore reinforced by its recognition by others as *his* product (Marx, 1844b, pp.121-122). Secondly in many instances the product itself might be the result of collective activity and thus the duplication could constitute an expression of social being. This latter will become important when we turn to discussion of the commodity.

We have said then, that 'development' and 'duplication' are two facets of the self-creation of the embodied subject through practical activity on a Marxian account. A third facet might be called 'extension'. As with Hegel, Marx in various places suggests that the agent in acting on the environment *extends into it*. He asserts for example that "Nature is the inorganic body of man" (Marx, 1844a, p.81) As with Hegel again, there is some overlap here with the idea of 'duplication', for seeing oneself reflected in the surrounding environment or expressing oneself in ones' products can both be taken as examples of agent extension. However Marx sometimes appears to mean more than this. One particularly significant example occurs in a discussion of the fashioning of tools from natural resources, where it is asserted that "nature becomes one of the organs of his [the productive individual's] activity, one that he annexes to his own bodily organs" (Marx, 1867, p.285.) In this passage a tool serves more literally as an extension of the agent. It enlarges the domain of activities that the agent is able to carry out and so becomes effectively a proxy body part. We might surmise also that Marx intended a subjective element to be implied here – that the tool is *experienced* as a spatial extension of the body – for although no explicit comment is made to this effect in the cited passage, it is clear from material elsewhere that he was as much concerned with

the subjective experience of the interface between man and technology as with its objective repercussions (see section below – 'The Machine')

The Material Basis of Thought

One upshot of the fact that the embodied subject 'creates' or 'realizes' himself through practical interaction with the material world is that the way that productive activity is organized in a society (the 'mode of production') will affect who s/he is:

> The mode of production must not be considered simply as being the production of the physical existence of the individuals. Rather it is a definite form of activity of these individuals, a definite form of expressing their life, a definite mode of life on their part. As individuals express their life, so they are. (Marx & Engels, 1970b, p.42)

Marx is partly just making what is almost a logical point here – from a third-person standpoint we are what we do, as a behaviourist might say, and therefore anything that impacts on what we do must also have an impact on who we are. However there is also a more cognitively significant implication. The manner in which production is organised affects not only who we are in a third person sense, but has repercussions for our subjective conception of the world. Marx expresses this idea in rather negative sounding terms as follows:

> The phantoms formed in the human brain are also, necessarily, sublimates of their material life process, which is empirically verifiable and bound to material premises. Life is not determined by consciousness but consciousness by life. (Marx & Engels, 1970b, p.47)

The 'phantoms' that Marx refers to here are the ideas that we have. We might separate out two fairly distinct, but related, arguments in this passage. Firstly there is the argument that the immediate subjective consciousness of each individual human being – the unreflective way he/she relates to the world - is affected by the type of productivity he engages in and more generally by his/her relationship to the world of production. What this might mean in the arena of modern industrial production we will look at shortly when we examine 'alienated labour'.   Secondly there is the suggestion that the way production is organized in a society, and the way that society is organized by production, has a determinate effect on what theoretical ideas come to dominate.

Certain outlooks, in the realms of philosophy or science for example, although they may make no explicit reference to particular forms of socio-economic organization, nevertheless reflect these forms in the assumptions they make. This idea is of course commonplace now – it serves as the starting point for 'Marxist' research in many areas of academia, particularly in social science and the humanities. One application of this idea from the realm of biology is Levins and Lewontin's (1985) suggestion that Darwinian evolutionary theory could not have got a foothold in a feudal society based on "fixed hereditary relations" because in such a social world an outlook that "made change an essential feature of natural systems" would have been inconceivable (pp.11-12). Only in a society in which productive relations had become revolutionized by capitalism could such ideas appear congenial. We will look further at the work of Levins and Lewontin in the next chapter.

Determinism

In highlighting how Marx thought that ideas are 'sublimates' of productive relations (and the societies they produce) the question arises of whether Marx's materialism is one that implies determinism. In fact there is some ambiguity here. It is often remarked that Marx rejected 'crude determinism' or 'mechanical materialism' (e.g. Engels, 1887 ; Levine, 1975) and his work seems in the most part to confirm this. In this respect his position is perhaps best summed up by his dictum that that "circumstances make men, just as much as men make circumstances" (Marx & Engels, 1970b, p.59). Reversing the emphasis here we might observe that Marx does not think that human activity is dictated by material conditions via a one way causal chain – like the movement of a ball on a snooker table after collision with another ball. Rather man is a "conscious active agent" (Levine, 1975, p.8) whose practice reciprocally influences and is influenced by those conditions. This is a perspective which recognizes the autonomous, self-creating nature of the subject whilst also acknowledging his embeddedness in the physical world.

Some commentators have seen in this perspective a potential threat to materialism insofar as equal weight appears to be attached to the realm of the self-determined agent as to the material environment. For this reason they have gone on to posit a material

'first cause' as a means of emphasizing the priority of physical circumstances. Thus Arthur (1970):

> Any Marxian attempt to resolve the apparent antithesis between mechanical determination and self conscious activity must include the point that in the first instance material circumstances condition us, however much we revolutionise those conditions later. (p.25)

Such speculation, however, seems a little misconceived for it is difficult to see *how* 'material circumstances' if they can 'condition us' once (by which Arthur presumably means 'determine us') should fail to do so in an ongoing fashion. Arthur is here perhaps being 'insufficiently dialectical' in searching for an initiating material event which gets the ball rolling in standard linear causal fashion rather than accepting the inherently reciprocal nature of the relationship between agent and material world on a Marxian account. This is not to suggest, however, that it is an easy matter characterizing circular causal relationships such that questions of temporal priority cease to be a concern. We will return to this topic when we look at Merleau-Ponty in the next chapter.

Nevertheless, whilst Marx does stress the autonomy of the self-creating subject, there are aspects of his work which are at least more amenable to a 'deterministic' interpretation. Some of his writings on historical development, for example, have been taken as implying, in Hegelian fashion, a degree of inevitability in the progression from one mode of social organization to another. (We will see an example of this when we look at the 'negation of the negation' in the next chapter.) It is possible to see in such inevitability an historical determinist thesis which places limitations on the effective agency of human beings, as well as justification for placing limits on their actual freedom (see e.g. Popper, 1945.) We will not discuss this complex issue further here. We will just stress that insofar as there are 'two Marx's' with respect to this question - a non-determinist and a possibly more determinist one (Humphries, 1971) - the perspective and sympathies of this thesis lie firmly with the former.

Marx contra Hegel on Alienation

We said earlier that on a Marxian analysis man is an essentially productive being, but that how he produces has a determinate effect on who he is and how he relates to the

world. It follows from this that terms like 'production' or 'labour', when presented in the abstract, do not sufficiently describe their likely effects on the agent. Productive activity, unqualified, might be a good or a bad thing for a human being. It is here that another area of disagreement between Hegel and Marx emerges. We saw in Hegel's Lord and Bondsman story, that whilst the Bondsman initially experiences his productive activity and its object as something negative and alienating, over time it becomes a source of empowerment and affirmation. The bondsman both acquires a "mind of his own" (Hegel, 1807, p.119) and appropriates the world as part of himself.

The problem, from a Marxian point of view, with this analysis is that the self-creation and affirmation occurs whilst the bondsman is still enslaved to the Lord, as though the latter fact will have no ultimate effect on the bondsman's development or relationship to the world. Alienation is surmounted through perseverance and an accompanying change of perspective on the part of the bondsman, without any accompanying change in social relations. As Rees puts it :

> Only the bondsman's consciousness has been transformed, not his real relation to the lord. There has been a revolution in thought but no revolution in social relations. (Rees, 1988, p.39)

Hegel is able to make this move because for him the distinction between alienated and unalienated labour is not tied to any particular material or productive arrangements but rather has to do with 'externality' per se (Lukacs, 1975 ; Rees, 1988 – although this is to some degree disputed by Sayers, 2003). Although many of the details of the individual accounts found in the *Phenomenology* and elsewhere – such as the earlier part of the Lord and Bondsman story, his account of the relation between individual and state, and the story of the Unhappy Consciousness – indicate that Hegel had an awareness of the particular concrete determinants of alienated consciousness, this is not his central concern. Rather the grand meta-narrative of the *Phenomenology* sees the 'thingness' of the objective world itself as problematic, as the externalization of self-consciousness and so as something to be "superseded and absorbed" (Marx, 1844a, p.103).

On Hegel's account, then, because alienation per se has no necessary link to any particular social configuration its supersession is not contingent on any change of social

relations. Marx's account is very different in that he sees alienation – subjectively, the perception of the world as a separate and alien power – as arising only from particular social/productive arrangements, and its supersession is therefore contingent on the practical abolition of those arrangements. It has nothing to do with the externality of the material world as such. Indeed Marx emphasizes that the 'thingness' of the world plays a central role in our development as embodied agents – "the fact that man is an embodied, living, real, sentient objective being means that he has real, sensuous objects as the objects of his life expression" (Marx, 1844a, p.104), and suggests also that our 'nature' is in this sense necessarily exterior to us – "A being that does not have its nature outside itself is not a natural being and has no part in the natural world" (ibid).

Free conscious activity and the 'whole man'

We have said that for Marx, unlike Hegel, alienation only arises in certain social/productive circumstances. Thus the individual's productive activy is either alienated or unalienated according to those circumstances – more specifically it is either 'free conscious activity' (Marx, 1844a, p.82) or 'alienated labour'. We will look at the latter shortly. As regards the former – it is difficult to summarise Marx's account of un-alienated labour in a fashion that is any more concise than Marx's own description of it in a passage from a text entitled *On James Mill*. We will therefore reproduce that passage in full below as it will make a useful reference point both in this chapter and in later ones:

> Let us suppose that we had carried out production as human beings. Each of us would have *in two ways affirmed* himself and the other person.
>
> 1) In my *production* I would have objectified my *individuality, its specific character*, and therefore enjoyed not only an individual *manifestation of my life* during the activity, but also when looking at the object I would have the individual pleasure of knowing my personality to be *objective, visible to the senses* and hence a power *beyond all doubt*.
>
> 2) In your enjoyment or use of my product I would have the *direct* enjoyment both of being conscious of having satisfied a *human* need by my work, that is, of having objectified *man's* essential nature, and of having thus created an object corresponding to the need of another *man's* essential nature.
>
> 3) I would have been for you the *mediator* between you and the species, and therefore would become recognised and felt by you yourself as a completion of your own

essential nature and as a necessary part of yourself, and consequently would know myself to be confirmed both in your thought and your love.

4) In the individual expression of my life I would have directly created your expression of your life, and therefore in my individual activity I would have directly *confirmed* and *realised* my true nature, my *human* nature, my *communal nature*.

Our products would be so many mirrors in which we saw reflected our essential nature. (Marx, 1844b, pp. 121-122)

We see, then, an all round sense of realization both of the expressive powers of the individual and of his, and others', social being. Like Hegel's spirit at the end of its journey, man is *at home* in himself and with the world. Unlike Hegel's account, however, there is nothing unearthly about this scenario- it is an 'at home-ness' firmly grounded in materiality. The elements of this tableau, and their configuration, could be characterized and enumerated in a number of different ways; but for our purposes we will say that there is a tripartite relationship between three components – the individual, his/her product and other individuals. This relationship is mediated by productive activity; intersubjective relations are facilitated by practical relations. The outcome might be described as a "oneness" (Fromm, 1966, p.31) or a "unity" (Fischer, 1968, p.15) between the elements involved – ultimately reducible to a unity between 'man' and 'nature' (Arthur, 1986, p.11).

Such a unity is sometimes characterized by Marx in more biological terms – for example as "the unity of living and active humanity with the natural inorganic conditions of their metabolic exchange with nature" (Marx, 1855, p.489), a point we shall return to in chapter 6. With respect to the individual it is also sometimes characterized in terms of the development or reconstitution of the "whole man" (Marx, 1844a, p.91; Fisher, 1968, pp.15-30). When activity is free the fragmentation imposed by industrial labour is superseded and the individual is once again complete insofar as s/he is unified with his/herself, with others and with the material world. The communal and practical unity which ushers in this era of the 'whole man', however, is only attainable when the collective activity in question is "voluntary" (Marx, 1844a, p.80) and "spontaneous" (ibid.) rather than coerced or enforced by necessity. Human beings realize themselves through practical activity but such activity must be allowed to unfold

in an unrestricted fashion, for the "free development of each is the condition for the free development of all." (Marx & Engels, 1848, p.26).

The idea of attaining a unity or 'wholeness' through practical activity becomes more meaningful once we contrast it with Marx's account of 'alienated labour'. We will now turn to this.

Alienated Labour

Marx uses the term 'alienated labour' to characterise the productive activity of human beings in modern industrial societies We should note that 'alienation' has both objective and subjective connotations in Marx's work (Sayers, 2012), referring both to the fact of separation - the loss of unity between subject and object - and to the subjective perception of that separation. The 'separation' itself is as multi faceted as the unity described in the passage from *On James Mill*. In the *Economic and Philosophic Manuscripts* Marx outlines four ways in which labour is alienated, although, as we shall see, it might be argued that not all of these 'ways' are completely distinct.

Alienation from the product of labour

The first dimension of alienated labour is 'alienation from the product of labour'. It is a bare material fact that the object upon which the industrial worker works is not in any sense his own. Rather it belongs to the company or individual who has purchased his labour.[2] The labourer is thus merely a means to an end for the Capitalist, a tool through which the Capitalist realizes his own projects. On this analysis the result of the labourer's activity, the product, is an expression of the Capitalist's own desires or requirements. Just as an artist might use a paintbrush to create a painting and still claim the result as his/her own so the Capitalist might use a labourer to produce a commodity and still claim the result as his own. As with the bondsman in the early part of Hegel's parable, the labourer is a mere extension of the Capitalist.

---

[2] Technically his 'labour power', but this need not concern us here.

Nevertheless the labourer has invested his own being into the object – for, as outlined above, that is the nature of productive activity. It is through productive activity that we become ourselves and, as Marx puts it, the result of that activity "is merely the summary of the activity of production" (Marx, 1844a, p.80). The individual realises himself through his activity and the product of his activity is thus an expression of his being. Such is the process of human self-creation, founded on "man's ability to create objects in which he realizes his subjectivity." (Avineri, 1970, p.85).

This combination of factors, that the labourer's product is not his own and that individuals realise themselves through productive activity, leaves the labourer in a rather uncomfortable position whereby the product of his labours appears to him as "an alien being, as a power independent of the producer."(Marx, 1844a, p.78) Marx elaborates on this point in terms of what we might call a feedback loop between man and the natural world. Whereas, when things are going well, as we suggested earlier, this loop acts as a catalyst for the development of human powers, when things are out of kilter the opposite effect can result:

> The more the worker externalizes himself in his work, the more powerful becomes the alien, objective world that he creates opposite himself, the poorer he becomes in his inner life, and the less he can call his own…The greater this activity, the more the worker is without an object. What the product of his labour is, that he is not. So the greater this product the less he is himself. The externalization of the worker in his product implies not only that his labour becomes an object, an exterior existence but also that it exists outside him, independent and alien, and becomes a self sufficient power opposite him, that the life that he has lent to the object affronts him, hostile and alien. (Marx, 1844a, p.79)

Alienation from one's own productive activity

We get a greater understanding of why this should be the case if we move on to the account given by Marx of the second dimension of alienated labour – the alienation of the worker from his own productive activity. Just as the end result of the worker's activity does not belong to him, so that activity itself does not belong to him. If production was carried out in a 'human' way, according to Marx it would develop naturally out of the spontaneous (individual and collective) desires and needs of human beings; it would be a means of fulfilling those desires and needs, both in itself and

through whatever change to the material world resulted. Production in Capitalist societies, however, cannot meet such criteria because work is organised in conformity to the needs of the Capitalist class and in accordance with their blueprint. Here there are two points, firstly that there are features of the way work is organized under Capitalism which are held to be at odds with human fulfilment – a point we shall return to later – and secondly that the very fact that the content of productive activity is pre organised and imposed upon the individual labourer means that it cannot be in accord with the worker's "own spontaneous activity". (Marx, 1844a, p.80)

The result of this imposition of a way of organizing productive activity is that "labour is exterior to the worker, that is, it does not belong to his essence. Therefore he does not confirm himself in his work" (ibid). Marx again goes on to describe this situation in terms of what might be called 'reversals'. The worker "is at home when he is not working, and when he works he is not at home" (ibid). Whereas, in a free society, work itself would be a means of self fulfilment and desirable in its own right, under capitalism "it is not the satisfaction of a need, but only a means to satisfy needs outside itself" (ibid). Marx, however, goes even further than this. Because Man is 'homo faber' (Avineri, 1970), because his essence is productive activity, so an antipathy towards the content and result of one's own activity can be said to result in an alienation from self:

> This relationship is the relationship of the worker to his own activity as something that is alien and does not belong to him; it is activity that it passivity, power that is weakness, procreation that is castration, the worker's own physical and intellectual energy, his personal life (for what is life except activity?) as an activity directed against himself, independent of him and not belonging to him. It is self-alienation, as above it was the alienation of the object. (Marx, 1844a, p.81)

Alienation of man from man

Alienation from the product of labour, and alienation from one's own productive activity ('self alienation') are, then, two aspects of alienated labour presented in Marx's *Economic and Philosophic Manuscripts*. A third aspect is man's alienation from his fellow men. We saw in the passage from *On James Mill* that productive activity, when undertaken freely, constitutes, amongst other things, the affirmation of the individual as a social being. We have also seen that if productive activity is not undertaken freely

then it is not the affirmation of the individual at all because neither the activity itself nor the object it produces belong to that individual. Thus alienation from the product of labour brings with it a third type of alienation – for if the object produced is not an expression of the power of the productive individual then it must be an expression of someone else's power, that of its owner – "if he [the producer] relates to the product of his labour, his objectified labour, as to an object that is alien, hostile, powerful and independent of him, this relationship implies that another man is the alien, hostile powerful and independent master of this object." (Marx, 1844a, p.83) Hence implicit in the alienation of the individual from the object he produces is the alienation of man from man. This is because, from the point of the productive individual, "the alien being to whom labour and the product of labour belongs, whom the labour serves and who enjoys its product, can only be man himself." (Marx, 1844a. p.84).

One might think here that such an account would hardly give us 'alienation of man from man' on a generalized basis. After all, all that we have described is the alienation of individual productive agents from those who own or control the means of production. Within that mass of individuals, it might be suggested, intersubjective relations could continue relatively uncontaminated. However it is perhaps no accident that Marx chose to describe this set up as 'alienation of man from man', rather than, for example, as 'alienation of productive individuals from the owners of the means of production', for the suggestion seems to be that it is the relationships one develops through productive activity that determine one's relationship to one's fellow men. This is perhaps one meaning of Marx's comment that "in the situation of alienated labour each man measures his relationship to other men by the relationship in which he finds himself placed as a worker." (Marx, 1844a, p.83) We might speculate too that Marx intended this also to apply to 'man' in the abstract - that it is primarily in the arena of productive activity that the relationship of individual men to 'man' in the general is developed (similar points are made by Meszaros, 1970; Ollman, 1971) On this reading we might perhaps say that the productive individual's fundamental experience of 'man' would be as the organising principle of his activity, and thus as something negative, antagonistic and 'other'.

Alienation from Species Being

Talk of alienation of the productive agent from 'man' in the abstract brings us to the fourth and final facet of alienated labour on Marx's account – alienation from 'species being'. Earlier we interpreted the term 'species being' as meaning something like Man's essence as a social and productive animal. On this interpretation 'alienation from species being' might be viewed as a 'cover all' term that combines all the other forms of alienation – alienation from the object of production, alienation from productive activity itself and alienation from other human beings – and one that does not add much else except the suggestion that the reunification of these elements corresponds with Man's essence. In fact, however, Marx (in 1844a) does in some places seem to use terms like 'species being', 'species life' and 'species characteristic' to suggest more than this, but there seems to be no clear consistency in meaning within these different uses. This is reflected in the diverse range of interpretations of 'species being' given by later commentators, ranging from "the essence of humanity" (Fromm, 1966, p.43) to "the ability of the species man to receive sensation" (Levine, 1975, p.3). Taken together these difficulties perhaps explain why some commentators have omitted mention of 'alienation from species being' altogether when enumerating the different aspects of alienated labour in Marx's work (e.g. Kamenka, 1969, pp. 19-20). We will not adopt that tactic here, but neither will we attempt a lengthy elaboration of the different uses of the terms 'species being', 'species life' etc. in order to understand what alienation from them might entail. Instead we will merely draw attention to a couple of passages which contain additional elements that will become relevant when we turn to look at the 'retreat from alienation' in cognitive science. The first passage is as follows:

> Man makes his vital activity itself into an object of his will and consciousness. He has a conscious vital activity. He is not immediately identical to any of his characterizations. Conscious vital activity differentiates man immediately from animal vital activity. It is this and this alone that makes man a species-being. (Marx, 1844a, p.82)

Here the suggestion seems to be that although, as essentially productive creatures, 'we are what we do' we are not, like other animals, to be identified with any one facet of our activity. As conscious creators we are able to learn and develop, and hence take on all manner of productive activity. In this sense, as Marx says, we produce in a 'universal' (ibid) fashion. Interpreting the passage in such a way one implication of

'alienation from species being' would be an enforced frustration of our potential to engage in a diverse range of productive activities. We will return to this theme when we look at the 'division of labour' in the next section.

The second passage suggests that also within the remit of our 'species life' is nature itself:

> Both with man and with animals the species-life consists physically in the fact that man (like animals) lives from inorganic nature, and the more universal man is than animals the more universal is the area of inorganic nature from which he lives. From the theoretical point of view, plants, animals, stones, air, light etc form part of human consciousness, partly as objects of natural science, partly as objects of art; they are his intellectual inorganic nature….From the practical point of view, too, they form a part of human life and activity. (Marx, 1844a, p.81)

We perhaps need not interpret too metaphysically Marx's intent when he says that nature 'forms part of human consciousness'. For purposes of this thesis at least the point is just that from a Marxian perspective, as well as being an essentially embodied and social animal, it is also part of his 'species being', his human essence, that Man is embedded in the natural world, a natural world which, as we have said, constitutes his "inorganic body" (Marx, 1844a, p.81). Moreover, as Man develops so he expands the range of natural resources through which he realizes himself (a point we will return to in chapter 5.)

Alienated labour is then presented as co-opting this relationship between Man and nature, turning his conscious vital activity against him and thereby alienating him from "his own body, nature exterior to him, and his intellectual being, his human essence." (Marx, 1844a, p.83) It thus reverses the natural (or at least 'desirable') order of things – that conscious productive activity through which Man in appropriate circumstances would realize himself as a social being becomes a mere means to an end. "Life itself appears merely as a means to life" (Marx, 1844a p.82) because, from the point of view of the individual, productive labour is used not to realize his/her humanity in a free and collective fashion but is undertaken simply to sustain his/her individual existence.

Division of Labour

Closely related, and to some degree implicit, in Marx's critique of alienated labour is his critique of the 'division of labour'. With this phrase Marx is not objecting to any differentiation of tasks per se, which might arise for a variety of reasons (Marx & Engels, 1970b, p.51). Rather his specific target is the bifurcation of productive activity which comes into being with the mental/manual division of labour. This is the scenario, as neatly encapsulated in Hegel's Lord and Bondsman parable, wherein the activity of one group of individuals is restricted to intellectual tasks whilst that of another is restricted to physical tasks.

Marx's objections are fourfold. The first is of least interest to us (in this thesis) as it relates to the inequality inherent in the mental/manual division of labour – differential relations to the production process imply that "enjoyment and labour, production and consumption – devolve on different individuals" (Marx & Engels, 1970b, p.52).

Of more interest to us is a second objection which links the division of labour to alienation. The work regime engendered by the division of labour is inherently alienating because it is not entered into voluntarily – "As activity is not voluntarily, but naturally, divided, man's own deed becomes an alien power opposed to him, which enslaves him instead of being controlled by him." (Marx & Engels, 1970b, p.54) Thus the division of labour is alienated, in part, because it does not flow from conscious choice but is imposed on the individual. We have already mentioned how alienated labour stands opposed to the "free development" (Marx & Engels, 1848, p.26) of the individual, and this account sits well with that conception of things. We shall ignore for now Marx's potentially problematic use of the term 'naturally', which appears to suggest an opposition between the 'voluntary' and the 'natural'. Although this raises interesting issues, there is not space to explore them here.

A third and connected objection to the division of labour is simply that it forces an "exclusive sphere of activity" (Marx & Engels, 1970b, p.54) onto the productive agent. The individual "is a hunter, a fisherman, a herdsman or a critical critic and must remain so if he does not want to lose his means of livelihood."(ibid). Marx contrasts this with a possible state of affairs in communist society:

> In communist society, where nobody has one exclusive sphere of activity but each can become accomplished in any branch he wishes, society regulates the general production and thus makes it possible for me to do one thing today and another tomorrow, to hunt in the morning, fish in the afternoon, rear cattle in the evening, criticise after dinner, just as I have a mind, without ever becoming hunter, fisherman, herdsman or critic. (Marx & Engels, 1970b, p.54)

The suggestion is, then, that the self-realization of the subject necessitates a transcendence of the narrow spheres of activity allowed within societies wherein a division of labour dominates.

We notice that here Marx's target is not restricted to the mental/manual division of labour but seems to be directed at the enforcement of any narrow specialism. Elsewhere Marx develops this idea particularly in relation to factory production insofar as it involves the breaking up of whole tasks into small component tasks. Enthusiasts for this 'detailed division of labour' (Braverman, 1974) included not only Adam Smith but also the 'father of computing' Charles Babbage who was one of the first to recognise that "dividing the craft cheapens the parts" (ibid, p.80), a principle that became known as "Babbage's great principle of economical production" (ibid). Boden (2006) points out in this respect that it was the architecture of the computing device which informed Babbage's ideas about the organization or production, rather than vice versa:

> [Babbage's work] discussed the organization of a factory or other large institution. Babbage said it had grown from the ideas he'd had while designing his 'calculating engine'. (Boden, 2006, pp.134-5)

For Marx, whose own account of the division of labour in factories was considerably influenced by Babbage (Wendling, 2009), the point, however, was that the fragmentation of productive activity to this degree brought with it the fragmentation of the subject who becomes "further dismembered" (Marx, 1846, p.136) as a result. We will return to this topic when we look at mechanized labour in the next section.

The fourth objection to the division of labour – specifically the mental/manual division of labour - is not of a kind with the first three, and is to some degree more of an

observation than an objection. Marx argues that the arrival of mental/manual division of labour facilitates the development of 'pure theory':

> From now on consciousness is in a position to emancipate itself from the world and to proceed to the formation of "pure" theory, theology, philosophy, ethics etc. (Marx & Engels, 1970b, p.52)

Here Marx is being simultaneously serious and ironic. He is being serious because he really does think that the separation of mental and manual tasks was a material prerequisite for the development of 'pure' theory. He is being ironic, however, because he is sceptical of the value of theory which has flown free from its material base, for this separation is reflected in the content of the theory itself which consequently is unable to grasp the world as it actually is. Hence the sense in which theory 'emancipates' itself by separating itself from the material world is questionable. With this latter point we can perhaps see echoes of Hegel's own critique of the "pure thought" of the stoic who detaches himself from the world. It will be remembered that in that instance the resultant ideas were dismissed as abstract and "lacking in the fullness of life" (Hegel, 1807, p.122). We might say that Marx sees the same result as arising on a systematic basis with the mental/manual division of labour. There is perhaps additional irony, however, in the fact that Marx's particular target here is German Idealism (Marx & Engels, 1970b, p.118) no better exemplar of which is there than Hegel – in whose work mind or spirit literally takes on a life of its own.

In outlining how the mental/manual division of labour devolves different forms of activity – in particular mental activity and physical activity – on different individuals, and engenders an ineffectual realm of pure thought in the process, we find ourselves back in the realm of the 'whole man'. The separation of thought and activity between individuals brings with it a conception of the individual agent as likewise compartmentalized. As Sohn-Rethel puts it :

> The class antagonism of capital and labour is linked intrinsically with the division of head and hand. (Sohn-Rethel, 1978, p.37)

The practical reunification of the agent - whereby mental and manual tasks are reintegrated in "free conscious activity" – is thus concomitant with a re-conceptualization of the agent as a unified being i.e. with the "unity of head and hand"

(ibid. p.139). The material reconstitution of the 'whole man' is in this way linked with a theoretical reconstitution of the agent whereby thought and practical activity are no longer considered in essential separation. We will explore this idea further when we look at 'praxis' later in this chapter.

The Machine

When we were looking at the division of labour we said that the 'detailed division of labour' - wherein tasks are broken down into component parts and each individual task is allotted to a separate individual – also has a fragmenting effect on the individual. One place where Marx makes remarks of this kind is in a more general discussion of 'methods for raising the social productivity of labour' (Marx, 1867, p.799). Such methods, he says,

> distort the worker into a fragment of a man, they degrade him to the level of an appendage of a machine. (ibid)

Because human beings realize themselves through productive activity, if that productive activity is then reduced to a series of repetitive movements then, so the argument goes, the producer himself is also reduced to a fragment of a human being.

However we also notice a further feature in the above quote, one which relates specifically to large scale mechanized labour. It is not just that it stultifies the experience of the productive individual by severely limiting his/her sphere of activity, it is that the individual becomes in some sense an extension of the machine. Marx expands on this train of thought in a separate passage where his object is not the individual worker but the mass of workers, represented as 'labour'. When production is mechanized, labour "appears merely to be a conscious organ, composed of individual living workers … it is itself only a limb of the system, whose unity exists not in the living workers but in the living (active) machinery" (McLellan, 1977, p.374)

We said earlier that, on Marx's account, human beings in various ways extend themselves into the environment through productive activity – that nature becomes their "inorganic body" (Marx, 1844a, p.81) and that tools become "annexed" as a kind of

proxy body part. Here, however a reversal seems to have taken place. In the case of the tool "the worker animated it with his own skill and activity" (McLellan, 1977, p.373) whereas here, as Marx says, "it is the machine which possesses skill and strength in place of the worker, is itself the virtuoso, with a soul of its own in the mechanical laws acting through it" (Marx, 1855, p.693)

A clear distinction has to be made, then, between the individual's relationship to the tool and his relationship (in the context of mechanized labour) to the machine. From the point of view of the worker at least, this is not a difference in scale but a difference in function: "In no respect is the machine the means of labour of the individual worker. Its distinctive character is not at all, as with the means of labour, that of transmitting the activity of the worker to its object." (McLellan, 1977, p.373). Rather the worker has himself become a means of transmitting the activity of the machine to the object.

We should note, however, that the machine itself has no purpose or knowledge of its own. That which it seems to have is derived from its creators and owners, (from whose perspective it *is* a kind of tool.) Marx points out that this serves as another source of alienation for the worker:

> The knowledge that obliges the inanimate parts of the machine, through their construction, to work appropriately as an automaton, does not exist in the consciousness of the worker, but acts upon him through the machine as an alien force, as the power of the machine itself. (McLellan, 1977, p.374)

Thus, we might say, the displaced intelligence of the 'masters' of the machine manifests itself as actual (and malevolent) machine intelligence to those whose activity it governs. Wendling (2009) talks in this connection of "machine fetishism", emphasizing how human qualities which the worker is unable to realize in his/her own practical activity are manifest in the operations of the machine:

> Machine fetishism is a product of technological alienation. Not only do workers use means of production that seem to operate by mystical and occult properties incomprehensible to the workers themselves, these machines increasingly display the very functions of which the worker is progressively deprived. (Wendling, 2009, p.57)

Wendling includes 'mobility, diversification of task and skill' (ibid.) amongst such functions, but it is 'intelligence' itself which is of more interest to us. We will return to this theme in chapter 4.

Commodity Fetishism

We saw in the earlier passage from *On James Mill* that for Marx the externalization of the agent in his product is an inherently social phenomenon because production itself is an inherently social phenomenon. Nevertheless, we might say, there is social and there is 'social'. Some types of product – e.g. physical tools – can be used in relative detachment from the social world, at least in principle. Other types of product have only a social existence. One such product is money. Insofar as money is money – rather than merely a round metal object, or a piece of paper, or pixels on a computer screen – it is entirely socially constituted. To reverse a point once made by Wittgenstein (1953, p.94), one cannot have private money (money for use by one person only) any more than one can have a private language (see also, Marx, 1855, p.84). Money, as a social, symbolic medium only exists in relation to a certain social practice and it is that social practice which determines what money is.

The social practice which defines what money is, is 'exchange' and in Marxian terminology the fact that it only has this role and no other practical attribute can be expressed by stating that it has 'exchange value' but no 'use value'. This is to be contrasted with the situation of commodities which have both 'use value' and 'exchange value' (Marx, 1867, p.126) – they can be exchanged (i.e. bought and sold) and also used in various ways. Having made this statement, however, we need to qualify it, for, at an economic meta-level, money is also a commodity (it can be bought and sold). Nevertheless the point remains that it is a special kind of commodity – one which functions as the 'universal equivalent' (Marx, 1867, p.181) of other commodities.

We will not here go into too much detail about Marx's analysis of 'exchange value'. We do however want to emphasize one important feature of Marx's account, namely

that money and commodities, as repositories of (exchange) value, are not what they appear to be. According to Marx, in the act of commodity exchange

> …the definite social relation between men themselves…assumes here, for them, the fantastic form of a relation between things. (Marx, 1867, p.165)

This is the idea known as "the fetishism of the commodity" (Marx, 1867, p.163) where Marx compares the relationship that human beings have to commodities with that which they once had with mythical gods. In both cases "the productions of the human brain appear as independent beings endowed with life, and entering into relationship both with one another and with the human race." (ibid) Here we find repeated many of those same elements which were present in the accounts of alienation outlined earlier. Once more that which is in fact a human product, an externalization of human properties, appears disguised as something 'other'. On this basis some commentators (e.g. Fromm, 1966, p.43) have viewed commodity fetishism as a straightforward example of the alienation of the worker from his/her product. However, it should be noted that with commodity fetishism it is not so much that the product is viewed as 'alien and hostile' (although this may be the case) but rather that the product appears to take on a life of its own. Marx gives the example of a table made out of wood. As a use value it is an "ordinary sensuous thing" (ibid), but viewed as a commodity it is transformed into something else:

> As soon as it emerges as a commodity, it changes into a thing which transcends sensuousness. It not only stands with its feet on the ground, but, in relation to all other commodities, it stands on its head, and evolves out of its wooden brain grotesque ideas, far more wonderful than if it were to begin dancing of its own free will. (Marx, 1867, pp.163-4)

What, then, is the sense in which commodities take on a life of their own?

From the point of view of production the suggestion seems to be that although human beings might be bound together physically by the production process, subjectively they remain atomized because their labour is not free or genuinely collective. "They do not know themselves as social beings", as Ollman puts it (Ollman, 1971, p.208). Thus their social being instead manifests itself in "mystical" (Marx, 1867, p.164) form through their products. These products take on the appearance of "social" things (ibid. p.165) because they are enmeshed in a particular social relationship (exchange).

Turning to exchange itself: If we imagine the simple act of buying and selling between two individuals we might say that in this act we have two human beings who stand in a concrete social relation to one another. However the logic which determines the unfolding of their interaction is determined by the abstract property of 'value' which they perceive as inherent in the objects that bring them together (and particularly in the money object, which has no other useful properties). Whilst 'value' is in fact a social relation between human beings (Ollman, 1971, p.195) it conceals itself as a relation between objects before which the individuals stand as passive spectators. Thus a reversal has taken place. Whereas, as outlined in *On James Mill*, when individuals produce freely their products exist as an extension and affirmation of their being, in the process of commodity exchange "the persons exist for one another merely as representatives… of commodities" (Marx, 1867, pp.178-9).

This is a difficult analysis and we almost certainly have not done it justice here. Nevertheless for our purposes the main point can be relatively simply summarized: In the case of commodity fetishism the externalization of the agent involves the (mistaken) attribution of human properties to human artifacts, a process which serves to conceal the real social origins of those properties.

Property

In the previous section we looked at the artifacts that are generated by the agent's productive activity from the point of view of exchange i.e. as commodities. We ought also to consider them, if only briefly, from the point of view of ownership i.e. as property.

We saw in chapter 1 that Hegel (1821) seems to have viewed property both as a 'right' of mankind and, in some passages at least, as on a continuum with labour as a means of self-realisation via appropriation of the external world. As regards the former, we are less concerned in this thesis with the political/ethical dimensions of property ownership from a Marxian perspective. Nevertheless we might note in this respect that Marx was not impressed in general with the language of 'rights' which he saw as applicable only

to "egoistic man, man separated from other men and the community" (Mclellan, 1977, p.52). Moreover, far from viewing private property as a human necessity or entitlement he saw it as internally related (Ollman, 1971) to other components of the capitalist labour process (division of labour, commodity production) and so as something that would be swept away by a communist revolution (Marx & Engels, 1848).

We need to be clear here however that the 'abolition of private property' for Marx was not about the forcible re appropriation of the personal possessions of individuals but was rather about the collective abolition of a system of property relations that allowed one individual to buy (or otherwise control) the labour of another:

> Communism deprives no man of the power to appropriate the products of society; all that it does is to deprive him of the power to subjugate the labour of others by means of such appropriations. (Marx & Engels, 1848, p.20)

In this sense it might be thought that Marx's position on property is not so antithetical to Hegel's as first appears, for Marx allows that an individual can 'appropriate' social products, and the passages from Hegel that we examined made reference only to the appropriation of personal property as an expression of will, personality etc…and not to the subjugation of others' labour. Nevertheless Marx would object to the central role property ownership plays in Hegel's account. It is not just that Hegel talks of 'property' in the abstract and so makes no reference to the extent of property ownership or how it was acquired, it is the idea that individual ownership of property is itself a means of realization for the individual that runs counter to Marx's own outlook. It is this aspect of his critique that is of more interest to us.

In Marx's account 'appropriation' of the world in its fullest sense – the "sensuous appropriation by and for man of human essence and human life" (Marx, 1844a, p.91) – is a multi-faceted activity. It involves, or can involve, the deployment of "all human senses and qualities" (ibid. p.92). Marx includes amongst these - "seeing, hearing, smell, tasting, feeling, thinking, contemplating, willing, acting, loving …"(ibid. p.91), which he also describes as "organs of man's individuality". It has, in addition, a social or communal dimension - in participating in social activity "I can appropriate the senses and enjoyment of other men." (ibid. p.92)

Here we are back with the "whole man" (ibid. p.91) i.e. the unified individual who is no longer alienated in the production process or reduced to a fragment of a human being by the division of labour. His realization is contingent upon the demise of Capitalism, which demise includes "the positive supersession of private property" (ibid. p.91). However, this is so not just because private property is internally related to other components of the capitalist order. Rather there is something limiting per se in the mode of human realization that private property permits. In the institution of private property:

> All physical and intellectual senses have been replaced by the simple alienation of all these senses, the sense of having. (Marx, 1844a, p.92)

Private property, says Marx, presents us with a particularly impoverished version of appropriation'. Mere 'having' is a static, lifeless orientation to the material world. It does not engage man's range of faculties as a complete, unified being, and by its nature it precludes the possibility of man realizing himself as a communal being. Thus, although there is nothing reprehensible in the fact of individuals having their own personal property, pace Hegel, property ownership should not be elevated as a means or model of self realization.

Praxis – the unity of thought and action

Finally we come to one of the portions of Marx's account that is most difficult to articulate, but at the same time one which will become particularly significant when we turn to modern day cognitive science. This is the notion of 'praxis'. We have said that on a Marxian analysis man, as 'Homo Faber', is an essentially practical creature. As Kitching puts it – in albeit a rather one-sided way – "if Marx had to pick an essential definition of human beings it would not be 'those creatures which think' but 'those creatures which act" (Kitching, 1988, p.27). Likewise we have also said that it is through practical activity that man creates himself, a dynamic which Arthur summarises as follows: "Man produces himself in and through this mediator [productive activity]; he develops new powers and needs in the dialectic of this practice" (Arthur, 1986, p.37). Expressed in world historical terms this means that

"…world history is nothing but the creation of man by human labour and the development of nature for man" (Marx, 1844a, p.95).

In emphasizing the centrality of practical activity to human existence we have already, to some degree, given an account of what Marx means by 'praxis'. However the term is commonly taken to have additional implications. Elaboration of such implications is somewhat hindered by the fact that the term seems to be associated with "a chaotic array of meanings" (Bernstein, 1999, p.79) in the work of Marxian commentators, and neither does it help that in Marx's own work "the terms 'praxis' and 'practice' have often been used interchangeably" (Jha, 2010, p.208). Nevertheless we might say that one key idea suggested by the term 'praxis' is that thought and action in some way form a "unity". 'Praxis' can thus be seen as another example of an Hegelian synthesis, a dialectical "unity of opposites" (Meszaros, 1970, p.17). Although the term itself has a long history (Bottomore, 1991, pp.435-440), it was first used in something approaching this fashion by the Hegelian theorist August Cieszkowski (1814-1894). Thus McLellan, for example, talks of the "synthesis of thought and action for which Cieszkowski coined the term...praxis" (McLellan, 1969, p.10). However it was Marx who took up and developed the idea in a thoroughgoing way. For him it was the antidote to the separation imposed on the individual by the division of labour. Earlier we made reference to the relation between the separation of thought from activity and the separation of mental from manual labour. This relation is well summarised in a passage by the communist theorist Pannekoek:

> In human labour, the material, physical side and the mental side are inseparable….Only because under capitalism the division of labour separated these two parts into functions of different classes….did intellectuals come to overlook their organic and social unity. (Pannekoek, 1937, p.448)

There is, then, the general idea of the inseparability of the mental and physical in practical activity. "All thought is inextricably part of human activity" (Kitching, 1988, p.28) and vice versa. It is only the division of labour which distorted our view of this fact.

Related to this general conception of a unity between thought and activity are some more specific theses which detail different senses in which they interpenetrate. Four of

these theses are outlined below. They are presented as elaborations of the central concept of 'praxis'; however it should be noted that this schematization is entirely our own. Other accounts of praxis might add elements that are here omitted, omit elements that are included here, or combine or separate elements that are here separated or combined.

1) Ideas are generated from practical activity

We saw in the section on 'the material basis of thought' that Marx referred to ideas as "sublimates" of a "material life process". (Marx & Engels, 1970b, p.47) In that discussion our emphasis was on the way in which the *content* of thought might be affected by this 'material life process'. Here our emphasis is more on the question of the direction of causation between ideas and material activity. Marx's main target in this respect is what he and many subsequent Marxists see as a defining feature of Idealism - the notion that human practical activity issues forth from the detached and privileged sphere of thinking, that 'ideas' are necessarily prior to activity, and activity is just the realization of ideas. As McLellan (1969) notes, such a conception was still common amongst Marx's Hegelian contemporaries and was particularly well encapsulated in Heine's maxim that "thought preceded action as lightning did thunder" (Heine, 1832, cited in McLellan, 1969, p.8). In place of this Marx wished to emphasize the priority of practical activity – thus:

> [materialism] does not explain practice from the idea but explains the formation of ideas from material practice. (Marx & Engels, 1970b, p.58)

There is a way of understanding this statement which might take one in a determinist direction. For example Pannekoek suggests that Marx's critique of Idealism is that it "explains the events of history as caused by the ideas of men" (Pannekoek, 1937, p.449). Whilst this undoubtedly *is* part of his critique (if articulated with the appropriate emphasis and clarified with additional explanation) it should not be taken to mean that, on a Marxian analysis, ideas themselves have no role to play in human practice. As Fromm says, Marxian materialism "does not pretend that ideas or ideals are not real or potent" (Fromm, 1966, p.19). Rather the point seems to be that thought does not bring action into being in a vacuum, but instead thoughts themselves are generated by the practical activity in which they are enmeshed. Clearly such an idea can be developed in

different ways. On one analysis we might still retain the ultimate separability of thought and action, but maintain that episodes of each are necessarily *interleaved* in purposeful activity. This seems to be the position of the nineteenth century Marxist Antonia Labriola (1897) who interpreted Marx's 'philosophy of praxis' as propounding a "unity of theory and practice where a cyclical movement of theory and practice will go on in a continuous pace."(as cited in Jha, 2010, p.208). Alternatively it may be that thought and activity form a unity at a more fundamental level. We will return to this topic when we look at 'epistemic action' in chapter 8.

2) Understanding human beings in their active relation to particular material circumstances facilitates answers to philosophical questions

Marx's position here is partly a response to various accounts of the human being which present 'Man' in the abstract. These include Feuerbach's account which, as we earlier suggested, influenced Marx's own adoption of the concepts 'species being', 'species life' etc.. Feuerbach (1841) had himself offered a critique of religious alienation which restored 'man' to the centre of the universe, but for Marx 'man' still remained an abstract category if considered in separation from the particularities of his active existence (Fromm, 1966, p.23; Meszaros, 1970, p.219). To understand what 'man' is one needs to look at the concrete active lives of real individuals:

> The first premises from which we begin are not arbitrary ones but real premises from which abstraction can only be made in the imagination. They are the real individuals, their activity and the material conditions under which they live, both those which they find already existing and those produced by their activity. (Marx & Engels, 1970b, p. 42)

> and later:

> Where speculation ends – in real life – there real positive science begins: the representation of the practical activity, of the practical development of men. (ibid. p.48)

We might observe, as others have also done (e.g. Rubinstein, 1981, p.125), that the tactic adopted by Marx here is somewhat similar to that adopted by the later Wittgenstein (1953). In both cases an examination of the concrete practices of individuals is recommended as an antidote to approaches which attempt to formulate explanatory concepts in the abstract – although in Wittgenstein's case the practices under examination are confined to linguistic ones. In fact the similarities between Marx

and Wittgenstein go much deeper than this, but here is not the place to examine them (see, Rubinstein, 1981; Easton, 1983; Kitching, 1994; Kitching & Pleasants, 2002.)

Once one has a concrete approach to the study of man as he actually is, in his active material life, the dissolution of certain philosophical puzzles follows:

> All mysteries which lead theory to mysticism find their rational solution in human practice and in the comprehension of this practice. (Marx, 1970a, p.122)

The 'mysticism' that Marx refers to is not (only) religious or spiritual mysticism but the 'mysticism' of philosophy itself which Marx conceives of as moving around in abstractions (in a bad Hegelian sense) and so as unable to deal with the concrete particularities of material life. More specifically, the puzzles which philosophy generates often arise as a result of a failure to take into account man's active relationship to the world. Instead it presents that relationship in terms of the passive contemplation of sense objects (Marx, 1970a, p.121). We will return to this topic in 4) below.

3) Practical activity itself provides a solution to philosophical questions

In the section above we said that, according to Marx, an understanding of man's practical relationship to the world could facilitate solutions to philosophical puzzles. However the citation we supplied suggested more than this. It is not just that "mysteries which lead theory to mysticism" (Marx, 1970a, p.122) find their solution in the *comprehension* of human practice, but that in some cases human practice itself provides a solution. There are perhaps two interpretations of this. On one interpretation, Marx is here recommending participation in practical activity as a means of facilitating solutions to philosophical problems. This seems to be Kitching's interpretation when he argues that:

> It is human activity which, as it were, 'joins' thought to the world. Conversely, it is speculating … in abstraction from practice, from activity, which creates nearly all philosophical puzzles." (Kitching, 1988, p.29)

This interpretation is perhaps similar to the position we outlined in 1) where we said that on a Marxian account of praxis, ideas are generated by practical activity. Here, we might say, the same idea is expressed as a recommendation rather than an observation.

There is however a stronger possible interpretation of Marx's intent. This interpretation is not necessarily in competition with that given above but rather might be seen as a supplementary thesis – the thesis that practical activity can itself sometimes *constitute* an answer to philosophical problems.

In some passages this has the appearance of a purely normative call to action, as though Marx were recommending the abandonment of thinking per se. Thus the famous eleventh thesis on Feuerbach: "The philosophers have only interpreted the world in various ways; the point is to change it."(Marx, 1970a, p.123) However underlying this seems to be both a critique of philosophical theory as an effective means of solving all of the problems that it poses, and the suggestion that such solutions might be *enacted* rather than articulated. This is particularly the case when the theoretical problems manifest themselves as an intractable dichotomy:

> It can be seen how the solution of theoretical opposition is only possible in a practical way, only through the practical energy of man, and their solution is thus by no means an exercise in epistemology but a real problem of life that philosophy could not solve just because it conceived of it as a purely theoretical task. (Marx, 1844a, p.93)

We have, then, the idea of a practical answer to a theoretical question. Resolution of certain theoretical conundrums is possible "only through practical means" (Fromm, 1966, p.29). We can perhaps detect in this notion faint echoes of Hegel's own manoeuvre in the *Logic*. It will be remembered that Hegel was accused of 'bringing movement into logic' because the transitions which he depicted (e.g. from 'being' to 'nothing' to 'becoming' – Hegel, 1817a, p.128) were temporal transitions arising from the mind's contemplation of the categories in question. He had thus, we might say, attempted to posit as a valid solution to the problems generated by a static logic (its inability to grasp relations in the real world), the action of the mind itself. Marx's strategy, we might say, is similar but replaces the 'mind' with real human beings and 'thought' with physical activity.

4) The world is brought forth through practical activity

We have said in several places that on a Marxian account, the individual subject creates himself as he alters the world. Reversing the order of priorities here we might perhaps also say that the subject does not submit to the world as he finds it but rather restructures it in accordance with his own needs, desires and perceptions (needs, desires and perceptions which are in turn transformed by this new material state of affairs.) There is nothing necessarily 'metaphysical' about this account – it is not a corollary of this position, for example, that - as normally construed in idealist terms – reality is a construct of the subject.

Nevertheless we should note that when Marx's topic is philosophy itself, his account of our active relationship to the material world, because it appropriates the language of philosophy, can appear to be putting forward a more 'metaphysical' position. One notable example of this is the first thesis on Feuerbach, where it is asserted that:

> The chief defect of all hitherto existing materialism (that of Feuerbach included) is that the thing, reality, sensuousness, is conceived only in the form of the object, or of contemplation, but not as sensuous human activity, practice... (Marx, 1970a, p.121)

As Marcuse has remarked, Marx's position here bears some similarities with Hegel's own rejection of 'sense certainty' as "the final criterion of the truth" (Marcuse, 1968, p.271). In both cases an analysis in terms of our active relationship to the world is recommended. Some have seen in this, and similar, passages evidence of a stronger thesis whereby the activity of Man determines the objectivity of the world. Kitching comes close to this when he remarks that:

> For Marx, it does not make sense to conceive of 'sense objects' as Feuerbach does, as if they were simply 'there' in reality. On the contrary they only become 'sense objects', objects known to human beings, when such beings actively appropriate them as part of their purposive life. (Kitching, 1988, p.28)

Rubinstein, influenced in part by Avineri (1970), goes even further than this, arguing that on Marx's account "reality is a product of the dialectical interaction of nature and human nature" and that therefore Marx has found a "middle ground between the materialist conception of the mind as a 'mirror' and the idealist reduction of the world to our idea of it." (Rubinstein, 1981, p.170)

We need to be careful here, however. As regards Kitching's point for example, whilst it is true, as Avineri argues, that "according to Marx, nature cannot be discussed as if it were severed from human action" (Avineri, 1970, p.70), it is not clear that Marx would have had any quibble with the idea that the material constituents of the world are "simply 'there' in reality" (Kitching, 1988, p.28). His concern, we might say, was more with a correct description of those material constituents in terms of our relationship to them. Some versions of materialism had typically presented the world as made up of statically conceived objects to which we stand as passive spectators. As suggested in 2) above, this gives rise to seemingly intractable philosophical puzzles centred on the separation of the subject and the object. These puzzles dissolve once one depicts the world and our relationship to it in a dynamic fashion. We will see an example of this when we come to look at the work of Alva Noë in chapter 5.

Similarly with Rubinstein's argument – whilst the notion that Marx's account occupies ground midway between materialism and idealism sounds pleasingly anti-dichotomous, it may not be true. Marx does not appear to have thought that "mind", conceived in the abstract, had any role in bringing forth "reality" (Rubinstein, 1981, p.170) conceived in the abstract. What he did believe was that real embodied individuals shaped the world in which they lived and were in turn shaped by it. This latter description seems to be entirely consistent with a materialist outlook.

Conclusion

In this chapter, then, we have looked at the philosophical portion of Marx's work, with particular emphasis on his conception of man as an embodied, practical agent who creates himself through productive activity. As a 'whole man' this agent has the potential for unification with his material and social environment through free conscious activity. However certain features of industrial societies – in particular the organisation and imposition of mechanized and fragmented labour for the purposes of commodity production – result in a disunified agent who is alienated from his own activity, from his product and from other human beings. In addition, the division of mental and manual labour gives rise to a disunified *conception* of the agent whereby

thought and action are considered always in separation. It is this disunity which Marx attempts to surmount via the concept of 'praxis'.

Many of theses themes have direct relevance to the field of cognitive science, but before we examine these connections we need to look briefly at some later Marxian theorists whose work also, in different ways, has a bearing on modern day cognitive science. This will be the topic of the next chapter.

Chapter 3 – 'The Bridge to Cognitive Science' - Engels and later Marxists

Introduction

In the following we look briefly at the work of six theorists who are to varying degrees Marxian in their outlook. One of these (Engels) has been selected mainly because his work develops and explains further some of the themes we have already highlighted in the work of Hegel and Marx, and which will become particularly significant when we turn to look at contemporary cognitive science. Four others (Vygotsky, Merleau-Ponty, Levins & Lewontin) are included because they have had an acknowledged influence on recent cognitive science. The final thinker (Alfred Sohn-Rethel) has not had any documented influence on cognitive science but his idiosyncratic interpretation of one aspect of Marx's analysis seems relevant to one particular issue in cognitive science (the 'online/offline' dichotomy – see chapter 5) The accounts given of these theorists are in no way meant as representational summaries of their work. Rather in each case we will only be looking at one or two key points from each that are relevant to the project at hand.

Friedrich Engels (1820 – 1895)

We have already made reference to Engels in the previous chapter because some of the works by Marx we cited were also co-authored by him. Engels was Marx's contemporary, friend and sponsor. Nevertheless, for better and worse, Engels has also made his own distinctive contributions to Marxist theory. McLellan summarises such contributions as follows – "Engels' contribution to the legacy of Marx culminated in the decisive role that he played in the tendency to transform Marx's views into a Weltanschauung, a philosophical system, an interpretation of the world." (McLellan, 1977, p.72). Reading between the lines, McLellan's suggestion here is that Engels'

contribution was a mixed blessing. Engels was a systematizer and tended to develop the ideas that existed in a more fluid, impressionistic form in Marx's own work, into rather rigid and dogmatically expressed formulae. This has both advantages and disadvantages. As regards the latter the biggest disadvantage has perhaps been for the world in general, for it was Engels' systematizing which helped pave the way for the ideology of 'dialectical materialism' (see e.g. Cornforth, 1952) beloved of Marxist dictatorships everywhere. For this reason some commentators have tried to disassociate Engels' outlook from Marx's completely. Levine, for instance, argues that a distinction should be made between Marxism and 'Engelsism' (Levine, 1975, p.xiv), and suggests that the success of soviet bolshevism was partly dependent on the presentation of a false unity between the two (ibid. p.xvi).

Nevertheless, Engels cannot entirely be blamed for the way his system was taken up. Moreover there is an advantage for us in that systematicity and rigidity brings with it a degree of explicitness. Whilst it is not exactly true, as we shall see, that Engels' arguments are always easy to follow – his language is less obscure than Hegel's, and less resonant with manifold meanings than Marx's.

Importantly, his work also contains some interesting and relevant insights, even if some of his scientific speculation misses the mark (Levins & Lewontin, 1985). In the following our focus will mainly be on Engels' contribution to 'dialectics', but before turning to this we will look very briefly at some of his remarks on 'human development' and 'planning'.

Human Development

Engels follows Marx in positing a reciprocal developmental relationship between man and nature. This is evidenced, for example, in a passage from the *Dialectics of Nature* where Engels asserts that:

> Natural science; like philosophy, has hitherto entirely neglected the influence of men's activity on their thought. It is precisely the alteration of nature by men, not solely nature as such, which is the most essential and immediate basis of human thought, and it is in the measure that man has learned to change nature that his intelligence has increased. (Engels, 1883, p.172)

We note however that the Hegelian notions of 'duplication', 'self-realization' etc…
which Marx developed in his own work have here been supplanted by the more
everyday notion of 'intelligence'. This is perhaps symptomatic of Engels' more
'scientific' orientation, for what seems to have interested Engels most was the
possibility of a scientifically plausible story detailing how a readily comprehensible
property such as 'intelligence' could be augmented through interaction with nature.
Although no such account is developed in any detail in Engels' work there are a few
attempts at preliminary sketches towards this end. In some such passages it is the role
of the hand which takes centre stage. The hand develops through human labour and
concomitant with this comes development of the human brain:

> Step by step with the development of the hand went that of the brain…the hand alone
> would never have achieved the steam engine if the brain had not attained a correlative
> development with it and parallel to it, and partly owing to it. (Engels, 1883, p.18)

Questions of plausibility or otherwise aside (although see e.g. Wilson, 1999 for a
summary of more recent hypotheses along these lines) this account is at least interesting
as an early description of a dynamic relationship between hand and brain acting to
facilitate cognitive development.

Planning

As a consequence of this development, human beings, on Engels' account become
further removed from other animals and are able to "make their own history
consciously" (Engels, 1883, p.18) This, for Engels entails planning. The actions of
conscious human beings become progressively less subject to "unforeseen effects and
uncontrolled forces" (ibid) because their enhanced cognitive capabilities make it more
likely that the results of their actions will "correspond to the aim laid down in advance"
(ibid). It is noteworthy that, having given in outline this description of a feature of
human cognition, Engels then makes a transition to a description of a political state of
affairs which he thinks corresponds to it – a mode of social organization based on
'planning':

Only conscious organisation of social production, in which production and distribution are carried on in a planned way, can lift mankind above the rest of the animal world as regards the social aspect… (Engels, 1883, p.19)

Although - taken at face value – the above description might be compatible with a number of modes of social organization, it is clear, particularly with the benefit of hindsight, that it gels well with some form of state socialism – i.e. a 'planned economy'. Thus we might say that, on Engels' analysis, planning agents and planned economies fit together. As implied in our introduction, the version of Marxism which underpins this thesis is one that runs counter to imposed state planning, and thus, to some extent, counter to Engels' position here. However Engels' suggestion that there is a link to be made between a conception of the agent as a 'planner' and the perceived necessity for imposed social planning is very relevant to our concerns. We will return to these points in the next chapter.

Dialectical Laws

We have said that Engels was more of a systematizer than Marx and was also more oriented to what he believed was a 'scientific' outlook. This is no more clear than in his treatment of dialectics. Although, as we have seen, a dialectical perspective informs Marx's work in a number of ways and Marx occasionally addresses the subject directly - for example when he is discussing the strengths and weaknesses of Hegel's account (Marx, 1844a, pp.96-109) - it is Engels who explicitly foregrounds dialectics as a subject in its own right (Ollman, 1971). It is at the centre of his major philosophical/scientific work *The Dialectics of Nature*, and also plays a significant role in another of his works, *Anti-Duhring*.

Engels argues that there are three laws of dialectics, laws which he claims that Hegel developed "in his idealist fashion" in the *Logic* (Engels, 1883, p.26). There is no ambiguity about the status of these laws, for Engels is quite clear that they are both derived from and apply to nature, society and human thought:

It is…from the history of nature and human society that the laws of dialectics are abstracted. For they are nothing but the most general laws of these two aspects of historical development, as well as of thought itself. (ibid)

The laws are given as:

1) The law of the transformation of quantity into quality and vice versa.
2) The law of the interpenetration of opposites.
3) The law of the negation of the negation.   (Engels, 1883, p.26)

We have seen aspects of these 'laws' already in our treatment of Hegel in chapter 1. The transformation of quantity into quality maps more or less directly on to Hegel's treatment of 'quantity', 'quality' and 'measure.' The interpenetration of opposites is perhaps implicit in Hegel's treatment of various antinomies such as 'cause' and 'effect'. The negation of the negation is a little more difficult to locate but, according to Engels underlies the progression in the *Logic* itself (ibid). To explain what Engels intended with each of these three laws it will help to look at some of his own examples.

The law of the transformation of quantity into quality

As we have said, this is familiar as it departs least from Hegel's own account. It is essentially the idea that qualitative changes take place on the basis of quantitative changes. However Engels attempts to couch his version in physics-like language. Thus: "For our purposes we could express this by saying that in nature, in a manner exactly fixed for each individual case, qualitative changes can only occur by the quantitative addition or subtraction of matter or motion (so-called energy)"(ibid). Engels takes one of Hegel's own examples as an illustration. As this is an example that will become significant when we later look at dynamic systems we will reproduce it here:

> …Thus for instance the temperature of water is first of all indifferent in relation to its state as a liquid; but by increasing or decreasing the temperature of liquid water a point is reached at which this state of cohesion alters and the water becomes transformed on the one side into steam and on the other into ice. (Engels, 1883, p.30)

The law of the interpenetration of opposites

This is somewhat more wide-ranging in Engels' account. It includes for example both the idea that two categorically opposed properties of an object or process are internally related to each other - e.g. that you can't have a positive charge without a negative

charge (Engels, 1883, p.163) - and the idea, already encountered in our summary of Hegel, of 'sublation', whereby what were considered two opposing properties are viewed as a unity denoted by a third term (Hegel's cause, effect and 'reciprocity' - Hegel, 1817a, pp.215-219). It also seems to be linked with the idea of simultaneous 'identity and difference'. Thus, in the case of organic beings Engels says –

> Every organic being is every moment the same and not the same; every moment it assimilates matter supplied from without, and gets rid of other matter; every moment some cells of its body die and others build themselves anew… (Engels, 1887, p.35)

Here the antinomous properties of the object under question arise from the fact that it is in flux. Engels does not think that a description in terms of (what would today be called) discrete state transitions (see chapter 4) adequately captures the matter, for the fluid transitional nature of the object is part of its being (an idea derived from Hegel's concept of 'becoming'). It is what it is by virtue of the fact that it is continually changing over time. It can therefore be justifiably be described as 'the same and not the same' at any particular 'moment'. We will not investigate the logic of this argument here.

The law of the negation of the negation

This expresses the idea that a thing (an organic being, a social system, thought processes etc.) may move through a set of transitions, each of which can be conceived as 'negating' the previous one, with the last transition restoring the thing to its original mode of being but at a "higher level of development" (Cornforth, 1952, p.212).

We might turn back to Marx for a well known example of the 'negation of the negation'. In the first volume of *Capital* Marx describes the process that (he thinks) precedes the arrival of communism. The starting point is an economic organisation of society wherein artisans and land owning peasants have some of their own personal property and a relative degree of freedom in using them – "the worker is the free proprietor of the conditions of labour, and sets them in motion himself" (Marx, 1867, p.927). Despite the relative freedom this brings it becomes, after progressing to a certain point, restricting for human (economic) development. Production is relatively isolated and the mode of social organizations means that, although the material

technology is available to produce on a grander scale this potential is never realized. Thus "at a certain stage of development, it brings into the world the material means of its own destruction" (ibid. p.928). It is at this stage that the private property of individuals (land, tools etc.) is "supplanted by capitalist private property" (ibid). The self-employed worker is no longer the owner of his own individual means of production but has these means expropriated and is forced, by circumstance, to work en masse for the benefit of the capitalist. This, according to Marx, is the first act of 'negation'. However Capitalism too brings with it the seeds of its own destruction. After a certain point the private ownership of the means of production by individual capitalists becomes a "fetter" on the further development of production by the centralized mass of workers who have been brought together by this mode of economic organisation. It is time for the 'expropriators to be expropriated'. (ibid. p.929). This, says Marx,

> is the negation of the negation. It does not re-establish private property, but it does indeed establish individual property on the basis of the achievements of the capitalist era: namely cooperation and the possession in common of the land and the means of production produced by labour itself. (Marx, 1867, p.929)

Engels spends many pages defending the 'negation of negation', particularly as represented in Marx's account, as it had come under attack from the philosopher Eugen Dühring (1833 -1921). The latter had accused Marx of using "a stupid analogy invented by Hegel, borrowed from the sphere of religion and based on the story of the fall of man and his redemption" (Engels, 1887, p.195) In Engels' counter attack examples from a wide range of areas are brought together, with varying degrees of success. One such example is the negation of 'primitive materialism' by idealism, followed by the negation of idealism by 'modern materialism' (ibid. p.190). Another, in part derived from Hegel (Rees, 1998, p.46), concerns the process whereby a grain of barley is 'negated' by its development into a plant, which is in turn 'negated' by the appearance of many grains of barley, and hence the return of the grain of barley at a (quantitatively) 'higher level' (Engels, 1887, p.187). We might also note a further example involving mathematics:

> Let us take any algebraic quantity whatever: for example, a. If this is negated, we get $-a$ (minus a). If we negate that negation, by multiplying $-a$ by $-a$, we get $+a^2$, i.e. the original positive quantity, but at a higher degree, raised to its second power. (Engels, 1887, p.188)

This example seems questionable. It is not clear, for instance, in what sense a negated number is further 'negated' by multiplication. Nevertheless there is something of relevance to cognitive science in the idea that dialectical processes can be thought to be at work in arithmetical operations. We shall come back to this point shortly.

Other features of Engels' dialectical outlook

As well as the three laws of dialectics Engels makes reference to two other features of a dialectical outlook which, although possibly derivable from these laws are perhaps best outlined explicitly.

Process and Movement

Engels emphasizes that a dialectical outlook is concerned with things in movement or as processes. Such an outlook, as he says, is derived from Hegel according to whom:

> the whole world, natural historical, intellectual is represented as a process i.e. as in constant motion, change, transformation, development; and the attempt is made to trace out the internal connection that makes a continuous whole of all this movement and development. (Engels, 1887, p. 37)

and, like Hegel, Engels also makes reference to Heraclitus' maxim that "everything is fluid, is constantly changing, constantly coming into being and passing away." (ibid. p.33)

For Engels, however, there is a need to materialize this outlook and give it a scientific appearance. He is thus primarily concerned with process and movement as properties of matter. To this end he asserts that "motion is the mode of existence of matter" (Engels, 1887, p.86) i.e. it is not just a contingent fact that matter moves but rather this is its essential characteristic. From this starting point he is able to suggest that it is matter in motion which eventually gives rise to consciousness. Such motion, he stresses, should not be conceived 'mechanically' in the way of primitive materialism but rather should be conceived dialectically, with quantitative changes facilitating the emergence of qualitatively new properties:

> The motion of matter is not merely crude mechanical motion, mere change of place, it is heat and light, electric and magnetic stress, chemical combination and dissociation, life and, finally, consciousness. (Engels, 1883, p.21)

Included in this critique of "mechanical materialism" was also the idea, derived from Hegel, that the shift from quantitative change to qualitative change is not uni-directional but reciprocal – qualitative change brings about quantitative change and vice versa. (Engels, 1887, p.321). Thus there is an overlap of two dialectical motifs – that of "reciprocal action" (ibid) and that of 'quantitative-qualitative co-dependence'. A dialectical, rather than mechanical, conception of processes incorporates both, according to Engels.

We should note that Engels was keen to find a mathematics that would be suitable for expressing such dialectical processes. We have already seen how he thought that the 'negation of the negation' could be manifest in rudimentary algebraic operations. In the case of process and movement he makes frequent reference to the 'differential calculus' and observes that "the differential calculus for the first time made it possible for natural science to represent *processes* mathematically and not only *states*: motion." (Engels, 1883, p.251) He also remarks that "in its operations with variable quantities mathematics itself enters the field of dialectics" (Engels, 1887, p.168). Although some have questioned Engels' own competence as a mathematician (Van Heijenoort, 1948) it is significant that here he is willing to countenance the expression of an all encompassing dialectical view of the world – one which he thinks applicable to myriad social, natural and cognitive processes – in terms of a differential calculus. We will return to this theme when we look at Levins and Lewontin later in this chapter, and also when we look at 'dynamics' in chapter 5.

Holism

Finally, as was the case with Hegel, holism is key to Engels' version of dialectics. In fact this feature rarely appears in separation from emphasis on processes and motion - thus, for example: "dialectics… comprehends things in their essential connection, concatenation, motion, origin and ending." (Engels, 1887, p.36) To get a better sense of what Engels means by 'understanding things in their essential connection' it is useful to look at the perspective against which Engels is pitching himself. This is the perspective

which he calls 'metaphysics' and which he identifies to some extent with the outlook of non dialectical science:

> [scientific method] has left us as us as legacy the habit of observing natural objects and processes in isolation, apart from their connection with the vast whole; of observing them in repose, not in motion, as constants not as essentially variables…To the metaphysician, things and their mental reflexes, ideas, are isolated, are to be considered one after the other and apart from each other, are objects of investigation, fixed, rigid, given once for all.   (Engels, 1887. p.34)

Here, then, Engels' conception of non-dialectical, 'metaphysical' thought is very similar to Hegel's conception of the 'Understanding' (also identified to some degree with the science of the time - Norman, 1976, p.39). In both cases non-dialectical thought is unable to understand its object in its essential movement and interconnectedness and so produces a static, fragmented and thus inadequate conception of this object.

Engels, again like Hegel, however also suggests that non-dialectical thought has its uses up to a point. Indeed he goes so far as to say that it is "justifiable and necessary" (Engels, 1887, p.35) in a number of domains. Thus for example cause and effect are held to be "good in their application to individual cases". (ibid) Nevertheless when one goes beyond the individual case and examines things in their "general connection with the universe as a whole" (ibid. p.36) such conceptions are found wanting. Metaphysical thought reaches its limit "beyond which it becomes one-sided, restricted, abstract, lost in insoluble contradictions" (ibid. p.35)

Lev Vygotsky (1896 – 1934)

From Engels we turn to Lev Vygotsky. The contrast between these two theorists is quite marked. Whereas Engels sought to incorporate all human, social and material processes within a grand dialectical schema, Vygotsky was more focussed, less ideological and perhaps more genuinely empirical in his approach. Vygotsky was a psychologist who began his career in the early years of the Soviet Union. Some have

seen in this fact the possibility that Vygotsky might have been a Marxist out of necessity rather than choice, and therefore that the Marxian elements of his account might be inessential decoration rather than core ingredients. Veresov for example, suggests that some of Vygotsky's major influences were not Marxist thinkers but thinkers from the 'silver age of Russian culture' (Veresov, 2005, p.32). However whilst it may be true that "it was not only Marxism which influenced Vygotsky" (ibid. p.45) it is clear that many of Vygotsky's ideas, as well as being explicitly presented as the product of a Marxian outlook, actually have content consistent with a keen knowledge of and interest in Hegelian and Marxian concerns. As Cole & Scribner put it "Vygotsky clearly viewed Marxist thought as a valuable scientific resource from very early in his career" (1978, p.6). Similarly, as Toulmin has said:

> Vygotsky was more than happy to call himself a Marxist…the general form provided by a 'historical materialist' philosophy gave him the basis he needed for developing an integrated account of the relations between developmental psychology and clinical neurology, cultural anthropology and the psychology of art…(Toulmin, 1978, no page ref)

In our short account we will not be investigating Vygotsky's output in all of these areas but instead we will look briefly at just two of his key ideas - those of 'psychological tools' and the 'social origin of higher mental functions'.

Psychological Tools

In a 1930 lecture Vygotsky notes that:

> In the behaviour of men we encounter quite a number of artificial devices for mastering his own mental processes…these devices can justifiably and conventionally be called psychological tools or instruments. (Vygotsky, 1930, p.1)

Vygotsky has a wide ranging conception of the sorts of things that can be called 'tools' in this sense – they include language itself as well as various numeric systems, mnemonic techniques, diagrams ands maps and "other conventional signs" (ibid.) He is clear that although the description of such items as 'tools' is an analogy, and therefore like all analogies has its limits, it is capable of being used with some precision in the sense that one can outline exactly what the commonalities and differences between standard ('technical') tools and psychological tools are. In particular, following Hegel,

Vygotsky thinks that tools of both kinds can be characterized as a 'middle term' (ibid. p.3) – a means of 'mediation' - between 'the activity of man' (ibid. p.4) and the object. However where psychological tools differ from technical tools is in the outcome of such mediation. Whereas technical tools are used with the intention of creating changes in the external physical world, psychological tools are meant to "act upon mind and behaviour" (ibid), facilitating "activity toward oneself, and not toward the object" (ibid). Thus, for example 'a knot in a handkerchief' (ibid. p.2) is used as an aid to memory rather than to (directly) bring about change in the external world.

The use of psychological tools, says Vygotsky, "enhances and immensely extends the possibilities of behaviour" (ibid. p.5) for it makes available to everyone an external resource by means of which increasingly sophisticated cognitive strategies can be deployed. Vygotsky sees in such a process confirmation of the Marxian thesis that "by acting on external nature and changing it [man] at the same time also changes his own nature…" (ibid p.4) This enabling function of psychological tools is often discussed under the heading of 'scaffolding' (e.g. Toulmin, (1978), Clark (1997)), a convention which we will adopt later in this thesis. However it perhaps should be made clear that Vygotsky himself did not use this term, and as others have pointed out (e.g. Shah-Shuja, 2008) the term was probably first used in a thoroughgoing Vygotskyan sense by Cazden (1979).

The social origin of higher mental functions

We have said already that language itself can be viewed as a psychological tool according to Vygotsky. It acts as such in a number of ways but Vygotsky seems to suggest that it is through 'internalization' that its role as a problem-solving tool becomes most significant:

> The greatest change in children's capacity to use language as a problem solving tool takes place … when socialized speech (which has previously been used to address an adult) is *turned inward*. Instead of appealing to the adult, children appeal to themselves. (Vygotsky, 1978, p. 27).

Vygotsky's argument is that verbal thought, or "inner speech" (Vygotsky, 1934, p.35) has its origins in social speech. Signs and words are initially used only for purposes of

social contact. Subsequently however children will then find themselves using speech when alone as a means of guiding and focusing their activity. This "egocentric speech" (ibid) then finally metamorphoses into verbal thought – which we might loosely call 'thinking' (ibid p.36). Hence what was originally a medium that facilitated social interaction (speech) has "become the basis of a new and superior form of activity" (Vygotsky, 1978, p.28).

In this description, then, the process whereby verbal thought comes into being is one that starts from the outside-in so to speak. As Vygotsky puts it "in our conception, the true direction of the development of thinking is not from the individual to the social, but from the social to the individual." (Vygotsky, 1934, p.36), an outlook in keeping with the Marxian maxim that "the relationship of man to himself first becomes objective to him through his relationship to other men" (Marx, 1844a, p.84). This is a frequent theme in Vygotsky's work applying not only to the transition from speech to verbal thought but to other operations as well, including 'voluntary attention', 'logical memory' and concept formation (Vygotsky, 1978, p.57). In all of these cases what at first functions as an interpersonal relation re-establishes itself as an "intrapersonal process" (ibid). We will return to this topic when we critically examine Andy Clark's conception of 'The Extended Mind' in chapter 7.

Maurice Merleau-Ponty (1908 – 1961)

Merleau-Ponty and Marxism

Our third Marxian theorist is the philosopher Maurice Merleau-Ponty. Merleau-Ponty has had a significant influence on the development of cognitive science. His work on embodiment is frequently cited not only by recent cognitive scientists (e.g. Varela et al., 1993; Clark, 1997; Noe, 2004; Gallagher, 2005; Thompson, 2007) but also by earlier critics of 'cognitivism' (most notably Dreyfus, 1972). What is less commonly acknowledged in the arena of cognitive science, however, is Merleau-Ponty's Marxist outlook.

In fact, as other commentators have pointed out (Miller, 1976; Potter, 2008), Merleau-Ponty's Marxism was not a fixed quantity but rather varied throughout his career. The (comparatively) younger Merleau-Ponty who wrote *Humanism and Terror* in 1947 was a committed 'orthodox' Marxist, at least in the sense that he was willing to countenance the possibility that Stalinist purges and show trials might be an historical necessity if it transpired that by such means one could "create a society without violence" (Merleau-Ponty, 1947, xxxvii). Such a position, we might remark, now seems only marginally more defensible than Heidegger's own support for national socialism a decade earlier (see e.g. Wolin, 1993). In Merleau-Ponty's case it was indicative of an outlook which, while it rejected any crude deterministic notions of an objective historical process separate from the self-conscious activity of human beings themselves (Merleau-Ponty, 1947, p.17) nevertheless thought, in Hegelian fashion, that there was a History to be realized, and in the light of which acts of state violence might come to be vindicated.

By 1955, with the publication of *Adventures of the Dialectic*, however Merleau-Ponty had become disillusioned with orthodox Marxism, and especially with an 'essentialist' conception of the proletariat (Miller, 1976, pp. 110-11) according to which those human beings who stood in a certain relation to the means of production had a predisposition to realize communism. As Miller (somewhat amusingly) expresses it:

> He came to question whether the proletariat in fact had a genuine commitment to radical social change. (1976, p.122)

Moreover, this later Merleau-Ponty began to see in Marxism a significant ambiguity on the question of human agency, which posited on one side a dialectic between subject and object but on the other a materialist determinism (ibid. p.125) expressing itself, for example, in the idea that there are "objective conditions" (Merleau-Ponty, 1955, p.208) which usher revolution onto the stage of world history. Miller views this change of perspective as the coming to fruition of a tension that had always existed between Merleau-Ponty's Marxism and his phenomenological work. Merleau-Ponty, he says, was "torn between phenomenology and a neo-Hegelian account of the meaning of History" (Miller, 1976, p.109). The former, according to Miller, placed its emphasis on

"the contingency and open-ended nature of meaning" (ibid. p.110) whereas an Hegelian-Marxian conception of history, though sometimes cloaked in the language of autonomy, leads in the direction of determinism.

Whilst there may be much in this analysis, we should not also ignore the possibility that aspects of Merleau-Ponty's phenomenological work may in fact have been *facilitated* by other – non-deterministic – elements of an Hegelian-Marxian outlook, elements, incidentally, which seem as evident in Merleau-Ponty's later work as his earlier. Below we will look briefly at examples of such 'elements' in two works that have since become relevant to cognitive science, *The Structure of Behaviour* (1942) and *The Phenomenology of Perception* (1945).

The Structure of Behaviour

Merleau-Ponty's work *The Structure of Behaviour* is notable for its dialectical character. Within cognitive science, Thompson (2007) has acknowledged that this is the case but he has not accredited the ultimate source of this perspective. In fact it becomes apparent that Merleau-Ponty has reproduced many of the dialectical themes found in Hegel and Engels. Thus for example Merleau-Ponty's conception of 'forms' is both holistic and dynamic, being that of "total processes whose properties are not the sum of those which the isolated parts would possess" (Merleau-Ponty, 1942, p.47) Moreover his description of the changes which 'physical form' undergoes in certain circumstances is familiar:

> It can happen that, submitted to external forces which increase and decrease in a continuous manner, the system, beyond a certain threshold, redistributes its own forces in a qualitatively different order, which is nevertheless only another expression of its immanent law. Thus, with form, a principle of discontinuity is introduced and the conditions for a development by leaps or crises, for an event or for a history, are given. (Merleau-Ponty, 1942, p.137)

Thompson, linking this description with autopoietic theory (see chapter 6), recognizes that the 'principle of discontinuity' forms part of a generalized 'dialectical' perspective in Merleau-Ponty's work (Thompson, 2007, pp.66-87). However he does not establish its provenance in Hegel's and Engels' work, or acknowledge that it has previously been

given a name – the 'law of the transformation of quantity into quality'. Neither does he appear to notice the Marxian overtones of the notion of an historical "development by leaps or crises".

Reciprocal causation

The Structure of Behaviour is also noteworthy for its references to 'circular causality', a key notion, as we have seen, in Hegelian and Marxian accounts. This is prominent, for example, in Merleau-Ponty's characterization of the relationship between the organism and its 'mileu'. In a section entitled 'The Human Order' (Merleau-Ponty, 1942, pp.160-184) we find strong echoes of Marx's and Engels' account of the reciprocal relationship between agent and material world, but with more explicit detail about the underlying causal story.

In this section Merleau-Ponty distinguishes his own dialectical account of the agent/environment relation from one which bases itself on "mechanical action" (ibid. p.160). In the case of mechanical action "the cause and effect are decomposable into real elements which have a one-to-one correspondence" (ibid). This is the view of 'cause' and 'effect' that Hegel had attributed to the Understanding and which Engels linked to 'metaphysical' (i.e. abstract, fragmented) reasoning. On this view of things "the cause is the necessary and sufficient condition of the effect" (ibid. p.161). Merleau-Ponty notes that even when one acknowledges some degree of reciprocal influence, within this framework any such reciprocity becomes decomposable into "a series of uni-directional determinations" (ibid), where, at any particular moment and in reference to any specifiable feature of either of the two relata, no feature can be playing the role of both cause and effect. Here we are perhaps reminded of how Labriola (1897) sought to unify 'theory' and 'practice' by interleaving them (see chapter 2.) In contrast, according to the dialectical conception of things, says Merleau-Ponty, "one cannot assign a moment in which the world acts on the organism, since the very effect of this 'action' expresses the internal law of the organism" (ibid).

The latter portion of this statement is a little difficult to interpret– for it is not immediately clear why temporal indeterminacy of the world's actions should follow

from the fact that their effects express an "internal law of the organism". As this is an important issue we might spend a little time discussing it here.

One interpretation (derived in part from Thompson, 2007, pp.69-70) seems to be that the actions of the world cannot be specified as 'inputs' in separation from the state of the organism itself. The features of the world that act on the organism are not things-in-themselves but rather, as causal objects, are only are what they are in relation to the organism. Such an idea recalls some of Hegel's own arguments for the unity of cause and effect – e.g. that certain features of the object affected facilitate the resulting effect - as well as the more general Marxian argument that "nature cannot be discussed as if it were severed from human action" (Avineri, 1970, p.70). We might note that in the case of Merleau-Ponty, one of the relevant features of this causal situation, expressive of the 'internal law of the organism', is that "physical stimuli act upon the organism only by eliciting a global response which will vary qualitatively when the stimuli vary quantitatively" (Merleau-Ponty, 1942, p.161) – hence it is the 'law of the transformation of quantity into quality' which comes into play as one determinant of the organism's response.

However, whilst such an argument might go some way towards clarifying what sort of thing an 'internal law of the organism' might be, and how it can be used as part of a description of a *material* unity between cause and effect, it is not clear how this need have a bearing on questions of *temporal* indeterminacy or unity. That an action or stimuli might bring about different effects according to the properties or state of the organism, (perhaps as modified by the history of previous interactions between world and organism) does not tell us that we cannot pinpoint in objective time when that stimuli occurred.

The same issue seems to hold even when we break down the boundaries between organism and mileu completely. Just as Marx thought that nature was the "inorganic body of man" (Marx, 1844a, p.81) so Merleau-Ponty appears to think that organism and mileu form a unity of some kind. This is suggested, for example, in his assertion that "the mutual exteriority of the organism and the mileu is surmounted along with the mutual exteriority of the stimuli" (Merleau-Ponty, 1942, p.161), as well as in the idea that the dependency between organism and mileu is so great that they "participate in the

same structure"(ibid). Again, however, this only seems to give us a material unity. It does not well articulate the notion of temporal indeterminacy. Things are perhaps better explained in an earlier passage from the book:

> When the eye and the ear follow an animal in flight, it is impossible to say "which started first" in the exchange of stimuli and responses. Since all the movements of the organism are always conditioned by external influences, once can if one wishes, treat behaviour as an effect of the mileu. But in the same way, since all stimulations which the organism receives have in turn been [made] possible only by its preceding movements which have culminated in exposing the receptor organ to the external influences one could also say that the behaviour is the first cause of all the stimulations. (Merleau-Ponty, 1942, p.13)

Like Engels, Merleau-Ponty is willing to concede that the partial view of a particular scientific approach - which in this instance is that which sees behaviour as a response to external stimuli – can have its uses. However in airing the converse view as part of an either/or dichotomy Merleau-Ponty's intention is the same as Hegel's in the *Logic*, i.e. to show that each option is as 'one-sided' as the other and that the true picture of things can only be attained by viewing the interchange between organism and mileu as a reciprocal unity. Events are "co-determinate all the way down" so to speak, and thus the standard order of precedence by means of which we normally differentiate 'cause' from 'effect' ceases to be relevant.

This perhaps stills leaves open the question of a possible distinction between not being able to differentiate what came first in a chain of events and actual simultaneity of events, but we will not pursue this here.

Phenomenology of Perception

Turning, then, to *The Phenomenology of Perception*, we will confine our attention to one passage from this work which has gained considerable attention in cognitive science (most recently, Clark, 2008; Rowlands, 2010) In this passage (Merleau-Ponty, 1945, pp.175-177) Merleau-Ponty describes the activity of "learning to find one's way among things with a stick" (ibid p.175). There are, for the purposes of this thesis, three points of interest in this account. Firstly there is the general point that for Merleau-Ponty such an activity is simultaneously a motor habit and "equally an example of a

perceptual habit" (ibid). Without further elaboration this is already significant as a challenge to the type of dichotomous reasoning which would seek to keep the two terms apart – as such it can be seen as another example of an attack on the 'fragmented' perspective of the Understanding (Hegel) or metaphysics (Engels).

We might then move on to consider Merleau-Ponty's actual description of stick use. Here he notes that "once the stick has become a familiar instrument, the world of feelable things recedes and now begins, not at the outer skin of the hand, but at the end of the stick."(Merleau-Ponty, 1945, p.176). Thus the stick becomes an extension of the agent – "it is no longer an object perceived by the blind man, but an instrument with which he perceives it. It is a bodily auxiliary, an extension of the bodily synthesis." (ibid) At this point comparisons might typically be made with Heidegger's notion of equipment as 'ready to hand' (Heidegger, 1927). However an equally strong comparison can be made with the much earlier Marxian conception of externalization in tool use. According to this conception, as we noted earlier, "nature becomes one of the organs of his [the productive individual's] activity, one that he annexes to his own bodily organs" (McLellan, 1977, p.456). Given Merleau-Ponty's Marxian outlook it seems likely that this account is at least as relevant here.

Finally, we might note the general lesson that Merleau-Ponty draws from the stick story - which is that "in the gaze we have at our disposal a natural instrument analogous to the blind man's stick" (Merleau-Ponty, 1945, p.177). Here the stick has become part of an analogy for the active nature of perception itself. Perception is not a passive process but an active relation to the world which "questions, ranges over or dwells on" (ibid.) the objects in that world. This line of reasoning, as we shall see, has been taken up in recent cognitive science most notably by Alva Noë (2004). It is also seems to extend one of the themes in the Marxian notion of 'praxis'. Although as we saw in the previous chapter, the concept of praxis was mainly posited as a counter to the dichotomy between *thought* and action – certain of Marx's arguments have relevance also to the *perception* / action dichotomy. This is particularly the case with Marx's criticism of traditional materialism in the 'First Theses on Feuerbach' wherein "the thing, reality, sensuousness, is conceived only in the form of the object, or of contemplation, but not as sensuous human activity, practice..." (Marx, 1970a, p.121) Whilst this passage makes no concrete reference to perception, the notion that, to understand the world as it

actually is, is not to take a passive, contemplative stance towards the object but an active, practical one, seems to gel well with the notion that perception is an active process. Merleau-Ponty, of course, was well acquainted with the concept of 'praxis', making extensive reference to it in *Adventures of the Dialectic*, and in one passage defining it specifically in relation to the 'First Thesis on Feuerbach' (Merleau-Ponty, 1955, p.48). It would therefore seem reasonable to assume that his concept of 'active perception' was heavily influenced by his Marxian outlook. (We will return to the topic of 'active perception' when we look at Noë in chapter 5.)

Alfred Sohn-Rethel (1889 – 1990)

Our fourth Marxian theorist is Alfred Sohn-Rethel. As was noted earlier, Sohn-Rethel, unlike the other Marxist thinkers in this chapter, cannot be said to have had any influence on cognitive science, acknowledged or unacknowledged, direct or indirect. Nor can it be said that amongst current views in cognitive science there is anything which mirrors his views to any significant extent (although there are some interesting comparisons to be made with Lakoff and Johnson – see chapter 5). Nevertheless he is included here because he is explicitly concerned with a theory of cognition, and because certain elements of his analysis would seem to have something of relevance to say to modern day cognitive science .We will not be able to explore the rather dense and intricate arguments given in his major work (Sohn Rethel, 1978) in any great detail. Rather our concern is with a specific concept at its centre – namely that of 'real abstraction'. However, in order to get to this centre some summary of the arguments that lead up to it are necessary.

A stated aim of Sohn-Rethel's work is to give "the true historical explanation of the enigmatic 'cognitive faculties' of civilized man" (ibid. p.34). Sohn-Rethel expresses sympathy with Kant insofar as he agrees with him that there are conceptual pre-requisites for advanced cognition and that these cannot be derived empirically from the world i.e. they cannot be "traced to the physical and sensorial capacity of experience" (ibid. p.38). However he disagrees with Kant that such conceptual pre-requisites are to be located in an 'a priori' transcendental subject. Indeed he thinks that such a perspective is itself a product of the "division between head and hand" (ibid. p.37)

brought about by the mental/manual division of labour. Instead he suggests that the abstract tools which facilitate our advanced conceptual abilities can be traced to the concrete practice of commodity exchange, and as such are neither passively empiricist nor idealist in origin.

We saw in the previous chapter that for Marx a commodity has a two-fold existence. As a substantial object it can be considered in relation to its physical properties as a 'use value'. However considered only as a commodity per se it is an 'exchange value'. In this latter guise all physical properties are abstracted from the object – as Marx puts it: "Not an atom of matter enters into the objectivity of commodities as values; in this it is the direct opposite of the coarsely sensuous objectivity of commodities as physical objects" (Marx, 1867, p.138). Sohn-Rethel sees in this abstraction from concreteness in the commodity, the origins of conceptual abstraction in cognition.

At this point we should perhaps note a certain ambiguity in Sohn-Rethel's argument, for he seems to equivocate between talk of 'the concept of abstraction' and talk of 'abstract concepts'. Clearly these are intimately related and it might be argued that one cannot have the one without the other, but it is worth pointing out that there is a distinction to be made. Examples given by Sohn-Rethel seem to straddle both categories. Thus, for example, the Greek philosophical notion of 'abstract universals' is held to be to be dependent on the arrival of coinage, for in the coin we find 'value' embodied as "an immutable substance, substance over which time has no power, and which stands in antithetic contrast to any matter found in nature" (Sohn-Rethel, 1978, p.59). Likewise 'atomicity' (in Democritus for example) can be seen as derivable from the value relation. As a physical object the commodity is a solid thing with particular differences to other objects, but as commodities all objects are equally and 'infinitely' divisible in terms of their value. This divisibility of value is again materially instantiated in money, which "must be divisible in order to leave the commodities undivided" (ibid. p.53). Similarly, though perhaps less abstractly, 'solipsism' is seen as derived from the act of commodity exchange, for this transaction is one which reinforces the separateness of its participants through their exclusive interest in and claim to the commodities being exchanged – in exchange "the action is social, the minds are private" (ibid. p.29)

These concepts, then, along with many others (including 'Abstract time and Space', 'Substance and Accident', 'Abstract Movement' and 'Strict Causality') are all "derivatives from material being" (ibid. p.201). This is not the empiricist 'material being' of "external nature and the material world" but rather is "social being" (ibid.) as manifest in the social practice of exchange. Conceptual abstraction has its roots in the practical activity of exchange in accordance with the Marxian materialist principle whereby the ideas of individuals are "sublimates of their material life process" (Marx & Engels, 1970b, p.47).

It is at this juncture that 'real abstraction' enters the story for, says Sohn-Rethel, the exchange of commodities itself constitutes an 'abstraction in practice' which occurs outside of the conscious awareness of participants. Indeed its success is dependent on participants not focussing on the abstract element of their interaction: "Were the abstractness to catch their minds their action would cease to be exchange and the abstraction would not arise" (Sohn-Rethel, 1978, p.27). The participants, instead, are motivated by, and occupied with, the individual uses to which their respective commodities (money or object) will be put. Thus, according to Sohn-Rethel, in exchange 'abstraction' inheres in the practice itself – "it is the action of exchange and the action alone that is abstract" (ibid. p.26). This is the meaning of 'real abstraction'.

By use of this phrase Sohn-Rethel says that he hopes to "dispose of the idea that abstraction is the exclusive privilege of thought" (ibid. p.7). This description of his project is perhaps a little misleading, for it would seem to imply that Sohn-Rethel's concern is with establishing that abstraction can take different forms, conceptual or practical, whereas his actual concern is also to show that an order of priorities exists in the origin of abstraction. As with Vygotsky, he appears to believe that the phenomenon under scrutiny existed first in social practice and only later as a property of individual minds. Nevertheless, for the purpose of this thesis, our main interest is in the idea itself. Abstraction can be a feature of practical activity as well as contemplative activity. We will return to this idea in chapter 5.

Richard Lewontin (1929–present) and Richard Levins (1930–present)

Finally we turn to Richard Lewontin and Richard Levins. Lewontin and Levins (henceforth 'L&L') are biologists who take an explicitly Marxist stance in several of their works, and most notably in the book *The Dialectical Biologist* (1985). Although chronologically they are the furthest removed from Engels in this chapter, it is they who are closest to him in spirit, offering what might be seen as an updated 'dialectics of nature'. Their work has had some significant influence on cognitive science, particularly on Varela, Thompson and Rosch (1993), although the Marxian nature of this influence is, with one honourable exception (Faith, 2000), rarely acknowledged in the field. In the following we shall try to summarise a few of L&L's key points.

Against the Metaphor of Adaptation

One of L&L's main arguments is directed against the Darwinian "metaphor of adaptation" (Levins & Lewontin, 1985, p.4). According to this view, say L&L:

> Organisms adapt to a changing external world, which poses problems that the organisms solve through evolution. (ibid)

L&L suggest that such a perspective places the organism in too passive a role. Organism and environment are presented as having "separate existences, separate properties". As a result the former becomes an "alienated object of external forces." (ibid) We might note that L&L's use of the term 'alienated' in this passage does not seem to imply the attribution of a subjective sense of estrangement to the organism (as in Hegel and Marx's accounts). Rather the suggestion is that the organism is 'alienated' in the objective sense that it has become separated, theoretically, from the network of dialectical relations that make it what it is. It is thus in this sense 'estranged' by an analysis which fails to see it in its actual connections with the world. (We will return to this point when we look at 'holism and emergence' below.) There is, however, the additional suggestion that such a theoretical manoeuvre is itself a reflection of subjectively experienced alienated social relations in the real world. As L&L note – "human beings see the natural world as a reflection of the social organisation that is the dominant reality of their lives" (ibid. p.12). Hence the depiction of the organism as a

plaything of external forces over which it has no control can be seen as an unconscious mirroring of the situation faced by human beings in their productive lives.

In place of an adaptationist perspective L&L wish to stress that "the organism is both the subject and object of evolution" (ibid. p.204). Implicit in this conception are a number of connected theses. These include that organisms do not merely find themselves in an environment but select that environment (ibid. p.53) and also that they modify the environment they find themselves in (ibid. p.54). This view of things recalls those aspects of Marxian theory which stress the active, productive nature of the human agent's relationship to the world; and L&L themselves trace their position back to Engels' own position on evolution:

> Engels captured the essential feature of human evolution; the very strong feedback between what people did and how they changed. He saw 'environment' not as a passive selective force external to the organism, but rather as the product of human activity. (ibid. p.253)

In some passages L&L go further than mere selection and modification, suggesting that the organism creates its environment *in toto*:

> It is impossible to avoid the conclusion that organisms construct every aspect of their environment themselves. They are not passive objects of external forces but the creators and modulators of these forces. The metaphor of adaptation must therefore be replaced by one of construction. (ibid. p.104).

Nevertheless, L&L do not intend to reverse the direction of causality as such but rather to suggest a reciprocal relation between organism and environment. As the organism modifies the environment so it, in turn, is modified by it. Thus organism and environment "actively co-determine each other" (ibid. p.89). In this respect L&L's account is not too dissimilar to Merleau-Ponty's account of the organism and its 'mileu', for L&L likewise characterize this relationship as one wherein "there is no longer a neat separation between cause (the environment) and effect (the organism)" (ibid. p.106). However whereas Merleau-Ponty's focus is on a lived reciprocity between subject and object, as evidenced for example in the 'animal in flight' story (Merleau-Ponty, 1942, p.13), L&L's focus is on reciprocity over evolutionary significant lengths of time, the reciprocity of a "co-evolutionary system."(ibid.).

Differential Equations and the end of Alienation

We have seen how L&L acknowledge their debt to Engels in their active account of evolutionary processes. They also follow Engels in a desire to express dialectical relations mathematically. Taking up his suggestion that "in its operations with variable quantities mathematics enters the field of dialectics" (Engels, 1887, p.168) L&L argue specifically that the reciprocal relation between organism and environment can be expressed through the use of differential equations:

> [The dialectical view which] breaks down the alienation between object-organism and subject-environment must be written as a pair of *coupled* differential equations. (L&L, 1985, p.105, authors' emphasis)

We will not here go into a detailed analysis of coupled differential equations. Suffice to say that they offer a mathematical means of describing the mutual interaction and co-evolution of two components (e.g. organism and environment) over time. The reciprocal nature of the interaction, and thus the internal relation between the two components, is captured in the fact that "the equation describing the evolution of each component contains a term that factors in the other system's current state" (Clark, 1998b, p.10).

What is particularly significant about L&L's account is that it centrally incorporates reference to the end of alienation in this context. To understand the reciprocal nature of the relation between organism and environment is to understand them in a non-alienated fashion i.e. in their essential unity rather than in separation from each other; and this unity can be expressed in mathematical terms. Thus the project which started with Hegel, the end of alienation through the union of subject and object, finds its mathematical culmination in a pair of coupled differential equations. We will return to this topic when we look at 'dynamics' in chapter 5.

Holism and Emergence

We have already alluded in various ways to the fact that, as with Hegel, Engels, and Merleau-Ponty, L&L's dialectical outlook is also 'holistic'. L&L's particular version of holism is pitched against what they call 'Cartesian Reductionism' (ibid. p.269), of

which they consider the Darwinian view to be an example (ibid. p.4). There are, according to L&L, four components of such a view:

1) There is a natural set of units or parts of which any whole system is made.
2) These units are homogenous within themselves, at least insofar as they affect the whole of which they are parts.
3) The parts are ontologically prior to the whole…
4) Causes are separate from effects…   (ibid. p.269)

As regards condition 3) this is taken as implying that parts exist in isolation, with their own intrinsic properties, but that they can come together to make wholes. L&L add that this model allows that in complex cases the interaction of the parts might produce additional properties for the whole. As regards 4), like Merleau-Ponty, L&L agree that this standard view allows for a two way influence between relata but that in such cases "there is no ambiguity about which is causing subject and which is caused object." (ibid)

We will not here investigate whether such a view is genuinely 'Cartesian', although we might perhaps remark that Descartes seems to be attributed an increasingly wide variety of views by critics, not all of which can be true of him. More significant, for our purposes is L&L's subsequent characterization of this view :

> We characterize the world described by these principles as the *alienated* world [author's italics], the world in which parts are separated from wholes and reified as things in themselves, causes separate from effects, subjects separated from objects. It is a physical world that mirrors the structure of the alienated social world in which it was conceived. (ibid. p.270).

Again, then, we have the use of the term 'alienated' to imply separation and fragmentation. An alienated worldview, on L&L's interpretation, is one which runs counter to holism by considering things in separation. We also note in this passage reference to the Marxian concept of 'reification'. This term, which Petrovic suggests denotes "a 'special case' of alienation" (Petrovic, 1991, p.463) is used by L&L to describe the way in which living relations become objectified as things. We will return to this term in chapter 7.

In contrast to the alienated worldview, say L&L, a holistic, dialectical world view is one that acknowledges that "part implies whole and whole implies part" (ibid. p.272).

L&L see in this reciprocity between part and whole, amongst other things, a different perspective on the concept of 'emergence' than is offered by standard accounts. It is not merely that new properties may arise from the interaction of the parts of a whole, which, as we saw earlier, was allowed by the traditional account, but that the properties of the parts themselves are also dependent on such a combination – "the properties [of the parts] come into existence in the interaction that makes the whole" (ibid. p.273).

Connectedly, another feature of L&L's dialectical holism is that, in contrast with 2) above, it does not attribute 'homogeneity' to component parts. Whilst "reductionist science" adopts such a strategy, from a dialectical point of view "things are assumed from the beginning to be internally heterogeneous at every level" (ibid. p.272). It is perhaps not quite clear what L&L's intention is with this remark. One might think that 'homogeneity' and 'heterogeneity' are relative terms anyway. A group of items are or are not homogenous according to certain criteria. There thus seems to be something a little 'one-sided' in positing a generalised internal heterogeneity per se. At the very least, we might say (somewhat flippantly) the parts must be homogenous in being equally heterogeneous. We will return to this topic in chapter 5.

An historical approach

We have said that L&L attack the fragmented, 'alienated' viewpoint of reductionist Darwinian science. We should add however that they temper this critique in one respect. Arguing from an 'historical' perspective they suggest that Darwinism was a necessary step in the development of science. Here the parallel is less with Engels' concession that Metaphysics has its uses up to a point, and more with the Marxian 'negation of the negation' in the history of forms of social organisation. Just as Capitalism replaced Feudalism and is in its turn replaced by Communism, in a sequence of necessary stages, so Darwinian reductionism replaced creationism, and is in its turn to be replaced by a dialectical approach (although in some places L&L present this as the 'completion' of Darwinism – ibid. p.106). In the case of Darwinian reductionism the idea seems to be that it was necessary to get a clear view of things in separation from each other before one could then work out their essential connections - "Only by alienating organism from environment and rigorously separating the

ontogenetic forces of natural selection could Darwin put evolutionary biology on the right track" (ibid).

Conclusion

In this chapter, then, we have looked at a number of Marxian theorists, whose work can collectively be thought of as constituting a bridge to cognitive science. Considered individually, in Engels' case the influence on cognitive science is an indirect one most evident in – though not restricted to – the dialectical naturalism of Levins and Lewontin; and in Sohn-Rethel's case, as we have said there is no chain of influence, direct or indirect on cognitive science, but his work is included both because it is already a theory of cognition, and because his concept of 'real abstraction' has something to say to recent cognitive science. In the case of Vygotsky, Merleau-Ponty and Levins & Lewontin, their work has had a direct and acknowledged influence on cognitive science.

In our next chapter we turn at last to cognitive science itself.

Chapter 4 – Alienation and Cognitive Science

Introduction

In the last three chapters we have spent quite some time looking at the work of Hegel, Marx, and Engels as well as at later Marxian thinkers. Our aim in doing so was to supply the theoretical resources, and to a small degree the historical justification, by means of which the current movement in cognitive science can be understood in a new light, as a 'retreat from alienation'. However before examining current cognitive science in terms of such a retreat, we need to spend some time looking at the object which engendered such a counter-movement, and at the question of why it might be thought to be alienated in the first place. This will be the aim of the current chapter – which will focus on what might loosely be called 'traditional' cognitive science.

A couple of points ought to be made clear at the outset. Firstly the account given of 'traditional' cognitive science is a) rather brief and b) presented largely from the point of view of its critics. In part mitigation we can only say that a), we have already spent a substantial portion of this work on summary rather than analysis, so a more fulsome account may not be appropriate at this juncture and b), the account given at least corresponds to the object as perceived by certain of its critics.

Secondly our breakdown in this chapter of different ways in which traditional cognitive science might be viewed as 'alienated' is not meant to be exhaustive. Nor are all elements of this account intended to correspond with counter elements in the hypothesized 'retreat'. In particular, as regards this latter point, some elements of the diagnosis in this chapter - e.g. the 'alienation from product' perceived as present in 'Strong AI' - have no principled anti thesis in the counter movement of recent cognitive science. Conversely, when we come to look at the positive theses of current cognitive science in the next chapter, we will get some additional insight into the sense in which the earlier conceptions of cognitive science might be thought 'alienated'.

Traditional Cognitive Science

It is no easy matter to try to summarise the key assumptions of 'traditional' cognitive science, for there are many complicating factors that need to be taken into account. First of all one would ideally like to avoid historical simplification. As Boden (2006) has pointed out not all cognitive scientists prior to a particular date seem to have held to all of the main tenets of a 'traditional' approach and conversely many of the perspectives now promulgated as part of a new wave of cognitive science were to some degree anticipated by the earliest AI researchers. Moreover if one wants to talk of a sustained critique of the core vales of traditional cognitive science within a framework conducive to the development of an 'embodied', 'situated' etc. perspective , then such an approach can be traced back to the sixties in the work of Hubert Dreyfus (1965, 1972).

There are also apparently terminological difficulties, which in fact relate to deeper theoretical issues about philosophy of cognitive science. It is convenient to use Haugeland's (1985) term 'GOFAI' (Good Old Fashioned Artificial Intelligence) to refer to those aspects of traditional cognitive science which have come under attack in recent years. The term is both less unwieldy than 'traditional cognitive science' and focuses our perspective more readily on certain key features of the classical approach (which is not to say that there is not debate about what 'GOFAI' means - see e.g. Chrisley, 2003, Boden, 2006). However in using the term we run the risk of suggesting either that our main interest is in AI research (a topic which I am barely qualified to discuss) or that we accept without question the assumption that the subject matter of AI is identical with that of psychology, neither of which is true.

In this chapter – and in the rest of this thesis - I will use the terms 'GOFAI' and 'traditional cognitive science' in a fairly unrigorous, sometimes overlapping way but will try to be explicit about the intended meaning in contexts where the distinction has significance.

In attempting to define traditional cognitive science we might begin with the relatively uncontroversial statement that it views mind either as a digital computer or as like a

digital computer in important respects. More specifically it views mind as a symbol processor. In this connection we note Newell and Simon's well known 'Physical Symbol System Hypothesis':

> A physical symbol system has necessary and sufficient means for general intelligent action. (Newell & Simon, 1985, p.87)

If we add that the symbols of a physical symbol system are "meaningless tokens"(ibid p.88; Harnad, 1990) processed on the basis of syntactic structure in accordance with explicit rules then it might theoretically be possible to end our definition of traditional cognitive science here; for on one view all of the other features of GOFAI are deducible from, or internally related to, this one, and traditional cognitive science, we might say, is just GOFAI applied across the board. Thus Haugeland, for instance, is quite happy to state that "GOFAI is a fairly coherent research tradition, based on a single basic idea: thinking as internal symbol manipulation" (Haugeland, 1996, p.25). However commentators will generally spell out other of its assumptions, whether or not they are derivable from this initial one (and differences of interpretation, e.g. on the question of where to place connectionism, suggest they are not so derivable.) How these assumptions might be described depends on whether the commentator is a critic or a friend, whether their concern is with AI or Cognitive Science (and relatedly whether they are concerned with machines or minds), and whether they are focussing on specifics or generalities.

One such assumption is of the temporal separation of 'thinking' processes from other processes. What these other processes are depends on the kind of agent we are talking about, but if we are talking about robots or human beings then prominent amongst these processes are perception and action. Although the quotation (above) from Newell and Simon might be read as suggesting that 'action' is at the root of their conception of intelligence, it is only so as an outcome. The symbol crunching which underpins intelligent thought on a GOFAI model takes place subsequent to any processing of perceptual input and prior to the symbol crunching which directly generates action. Hurley, talking generically about 'minds', has referred to this as the 'classical sandwich': "Perception and action are not just separate from one another, but also separate from the higher processes of cognition. The mind is a kind of sandwich, and

cognition is the filling" (Hurley, 2001, p.3). Torrance has described the temporal relationship between the last two components of this sandwich in an even more prosaic fashion as "working solutions to problems out in the intellect or in the head prior to putting them into practice in the 'real world'" (Torrance, 1999, p.53), a description perhaps pitched at the level of conscious processes. A more robotics-oriented description can be found in Brooks' notion of a "sense-model-plan-act framework" (Brooks, 1991a, p.570). Brooks recounts actual cases of robots designed in accordance with this architecture as follows: "They all sensed the world and tried to build two or three dimensional world models of it. Then, in each case, a planner could ignore the actual world, and operate in the model to produce a plan of action for the robot to achieve whatever goal it had been given" (ibid).

The temporal separation of 'thinking' processes (a description perhaps most problematic when applied to the Brooks example) from perception and action goes hand in hand with other sorts of separation. Again how one characterizes these other types of separation depends upon the precise nature of one's object and what one's priorities are. At the level of the processing unit(s) of a robot – or 'equivalently' at the level of the brain - the separation of 'thinking' processes from other process might be conceived of not only as temporal but as 'modular'. Thus Brooks describes how in a classical system "the perceptual modules deliver a symbolic description of the world and action modules take a symbolic description of desired actions and make sure they happen in the world. The central system, then is a symbolic information processor" (Brooks, 1991b, p.402). Typically the notion of 'modularity' implies (amongst many other things) that separate functional units have fixed locations (although these locations may be distributed across the system). Thus, for instance, Fodor claims that in the case of brains there is an "intimate association of modular systems with neural hardwiring" (Fodor, 1983, p.98).

At the level of the whole functioning (biological) system it is sometimes the complete brain that is regarded as the central processor and the locus of cognition. The 'separation' in this instance thus becomes that between brain and body – or if we are to keep up with the tripartite 'sandwich' metaphor (and if we are to replace processes with objects all the way through) – then it is between world (input), brain (processing) and body (output). An example along these lines is given by Clark when considering the

processes involved in catching a ball – "In the traditional model the brain takes in data, performs a complex computation that solves the problem (where will the ball land?) and then tells the body where to go" (Clark, 1999b, p.6).

At a higher level of abstraction the term 'brain' is often replaced by 'mind', and the traditional approach is described (more typically, although not always, by critics) as centrally involving a separation of mind from body. Iverson and Thelen, for example, argue that "from a cognitivist point of view the body is an output device that merely executes commands generated by symbol manipulation in the mind" (Iverson and Thelen, 1999, p.19). Once traditional cognitive science is conceived of in this way it is inevitable that one of its defining features is felt to be a Cartesian 'dualist' outlook (although the dualist metaphor perhaps does not sit easily with the tripartite 'sandwich' metaphor.) Characterizations in these terms are very common (e.g. Searle (1980); Gardner (1985); Varela (1993); Damasio (1994); Anderson (2003), Wheeler (2005)). However it should be noted that there are great differences, often acknowledged, in how this dualist heritage is interpreted[3].

In noting that traditional cognitive science is often taken as positing, for explanatory purposes, a separation between mind and body we ought also to note that it has often been regarded as having close affinities with a 'functionalist' outlook, and particularly with the notion of 'multiple realizability' (Putnam, 1967.) The descriptions of cognitive/computational processes it offers are such as to be divorced from any particular mode of physical instantiation. "The same algorithm may be implemented in quite different technologies", as Marr (1982, p.24) puts it. Thus, in the case of living beings, it is not necessary to have an understanding of physiological matters in order to have an understanding of cognition: "Cognitive processes can be understood in full abstraction from their physiological embodiment" (Gallagher, 2005, p.134).

Moving away from strictly architectural considerations, traditional cognitive science is also more generally characterized as emphasizing certain types of mentality. In the case of GOFAI, for example, these might be "higher level competences such as thought, reason, planning and problem solving" (Chrisley 2003, p.3). Here we have to separate

---

[3] ..which perhaps makes questionable the usefulness of talking about the Cartesian roots of Cognitive Science.

the question of what types of mindfulness students and practitioners of cognitive science have historically been concerned with (see e.g. Boden (2006) p.10-18) from the question of what types of mindfulness cognitive scientists have seen as most significant for agents. Although obviously closely related it is the latter of these two that is of relevance to us.

Once again characterizations of this sort vary in their level of generality and modes of individuation. Clark suggests that traditional cognitive science viewed "mind as a kind of logical reasoning device coupled with a store of explicit data – a kind of combination logic machine and filing cabinet" (Clark, 1997, p.1). Ignoring the storage/filing cabinet portion of this description for the time being, we might say that the 'logic machine' component has (at least) two implications. Firstly there is the idea that thinking is similar or identical to the kind of symbol manipulation described by Newell and Simon – a point we have already alluded to. Secondly, and quite separately, there is the further idea that the activity of the 'mind' as a whole can be reduced to such processes. All mentality is thinking[4], a conception of things which, as Torrance has pointed out, is "enshrined in the very name of Artificial Intelligence" (Torrance, 1999, p.48). Torrance refers to this second feature as 'intellectualism' (which he sees as complementary to 'computationalism') and describes it as that which "takes the ratiocinative, sequential, symbol-manipulating varieties of mental activity as the defining feature of mind" (ibid.)

A similar account is given by H&S Dreyfus, although they prefer to use the term 'cognitivism' – a term taken up by other commentators (e.g. Freeman & Nunez, (1999); Anderson (2003); Froese (2007)) but not always with the exact same connotations. For H&S cognitivism is the view that "all mental activity is cognitive – that perception, understanding, learning and action are all to be understood on the model of fact gathering, hypothesis formation, inference making and problem solving"(Dreyfus, 2002, p.1) a view which they see as the "culmination of the rationalist philosophical tradition" (ibid). On the Dreyfus' account in particular then, traditional cognitive science is portrayed not just as an approach which emphasizes higher order processes

---

[4] Here, in fact, we have skipped a stage – that "all thinking is logic like" – but these stages cannot be rigidly defined because their components – terms like 'cognition', 'thinking' etc – are themselves not rigidly defined.

but as one which seeks to understand all skilful activity as emanating from these processes.

So far we have given a brief characterization of 'traditional cognitive science' in terms of some of the features it held to be significant for both biological and artificial minds. In doing so we have at least implied a certain cross fertilization across domains. Strategies used to generate 'intelligent' output from artifacts were theorized as strategies also used by biological minds, with the latter being viewed as more 'computer-like' as a result. Conversely, assumptions about the intelligent action of biological agents (e.g. its emanation from disembodied, 'high-level' processes) were transposed to the artifactual world with the resultant artifacts being viewed as more 'mind-like' for that reason. In the latter case we ought to note how some adherents of GOFAI have not stopped short at positing the 'mind-likeness' of artifacts but have gone on to suggest that such artifacts have the potential to become genuine minds – or at least genuine intelligences.

This is the position that Searle (1980) calls 'Strong AI' – the idea that "the appropriately programmed computer really is a mind in the sense that computers given the right programs can be literally said to *understand* and have other cognitive states" (p.183 – author's italics). For some this assumption should be included as one of the central tenets of GOFAI. Boden (2006, p.702), for example, argues that a version of this assumption seems to be implicit in Newell and Simon's own position (see p.118), a physical symbol system having both "necessary and sufficient means for general intelligent action" (where the latter is taken to indicate "the same scope of intelligence as we see in human action" (Newell & Simon, p.87)).

However in noting this we ought also to note that there is perhaps some ambiguity in the notion of 'Strong AI'. Firstly there is an issue with the way that the phrase 'really is a mind' should be taken in Searle's account (and how similar phraseology should be taken in other accounts.) It is not clear that such a phrase itself is amenable to any one definitive interpretation, and neither does further elaboration along the same lines (e.g. that a computer can "be literally said to understand") necessarily help[5] That there are no

---

.[5] Here we are perhaps reminded of Wittgenstein's comment "It is as if I were to say: "You surely know what 'It is 5 o'clock here' means; so you also know what 'It's 5 o'clock on the sun' means".

accepted criteria for deciding whether a machine 'really' has a mind, despite various attempts from Turing (1950) onwards, leaves it open as to how such phrases are to be taken.

Secondly and connectedly, pace Boden, it also not clear that Newell & Simon's adherence to the belief that an artifact can exhibit "the same scope of intelligence as we see in human action" is the same as adherence to the belief that an "appropriately programmed computer really is a mind"; or, to take another example, it is not even clear that a later, stronger claim of Simon's – that "a computer simulation of thinking thinks" (Simon 1995, cited in Boden 1419) is identical to such a belief. Neither genuinely intelligent action nor genuine thinking need be identifiable with genuine 'mind'. Here there does at least seem to be some space for gradations of 'Strong AI'. (We might note in this connection, that Chrisley (1995) makes use of the terms 'Strong AI' and 'Weak Strong AI, although for different purposes.)

We will say no more, in an unqualified way, about traditional cognitive science, although further of its features will emerge in our assessment of it, to which we now turn.

Mind Thinking About Mind

In what sense or senses, then, could the above account be described as 'alienated'? As a starting point it might be useful to step back from the details of that account in order to make the general observation that the activity of cognitive science is one wherein mind thinks about mind, or, to be slightly more specific, wherein the minds of cognitive scientists think about mind. By itself this need not distinguish cognitive science from other endeavours such as psychoanalysis, meditation or even English Literature. However there is perhaps something distinctive in the way that cognitive science thinks *only* about mind - its object over all, whatever the particular practical purposes of individual projects and disciplines within cognitive science, is just the understanding of mind per se. Its declared interest is in mind 'in itself' rather than as a means to some other end. Whilst we may not want to make too much of this, we can perhaps already see Hegelian possibilities in such a description. For a thoroughly modern Hegelian,

who was also convinced of the central importance of cognitive science, it might be tempting to argue that cognitive science *is* self-conscious mind coming to know itself. Mind externalizes itself – makes itself into an object - in cognitive science, with the aim that it will eventually achieve unity through 'absolute knowledge' of itself. On this analysis Margaret Boden's recent two volume history of cognitive science (2006) could perhaps be seen as the contemporary equivalent of the *Phenomenology*.

Less fancifully, although we might not be convinced of the central role of cognitive science as the 'becoming' of human mind, there is perhaps a more acceptable interpretation of the above position which dispenses with metaphysics but allows that there might be a dialectic at work in cognitive science. In fact there seems to be more than one possible candidate for this role. We might suggest, for example, that cognitive science is dialectical in its *methodology*. We said in our brief summary of traditional cognitive science that there was a cross fertilization between the investigation of mind as a computer and the investigation of computers as mind. This relationship can be seen as dialectical both insofar as it is reciprocal and insofar as this reciprocity results in a 'unity' of sorts. The latter is evidenced to some extent by the deliberate synthesis of man and machine in the concept of the 'agent'. It is part of the strategy of cognitive science that it remains neutral in this respect so that commonalities can be established between human being and machine. We might observe, however, that when this neutrality in practice is taken as an identity in fact, the resultant world view could be construed as an alienating one. This is one of the points we hope to establish in this and later chapters.

More relevant to our Hegelian story above, however, is not this methodological dialectic, but the idea that cognitive science could be subject to an 'historical dialectic' (Arthur, 2004) - that certain dialectical patterns, if not 'laws', might be manifest in the transition from traditional cognitive science to more recent cognitive science. In order to explore this idea we need first to say something about the content of recent cognitive science. Thus we will postpone development of this train of thought until the next chapter.

Idealism

To return to the content of our earlier summary then, if cognitive science in general is 'mind thinking about mind', traditional cognitive science can be viewed as mind thinking about 'mind' in a narrowly mentalistic way i.e. as viewing all intelligent human behaviour as the outgrowth of 'cognition', considered in separation from any particular material manifestation or activity. In practice, if not in its manifestos, the essence of being human for a traditional cognitive scientist is being a thinker. Man is 'Homo Cogitans' rather than 'Homo Faber' (Ingold, 1987, p.62). Like the idealist philosophies that Marx denigrated, it might thus be seen as reproducing the assumptions of the mental/material division of labour – in particular the essential separation of 'head and hand' (Sohn-Rethel, 1978, p.37). Similarly, like Hegel's own idealist project, it is open to the criticism that it sets out from an alienated starting point by taking "the life process of the human brain i.e. thinking" and transforming it "into an independent subject" (Marx, 1873, p.14).

Of course traditional cognitive science does not in a literal sense propose that thought has an existence independent of the material world and so cannot be accused of metaphysical idealism. However in giving priority to 'the mind' in general, and to cognitive processes specifically, it does display some of the features which Marxists describe as 'idealist' in a weaker sense. In particular its emphasis on the *temporal* priority of cognition over action mirrors at the level of individual agency what Marxists attribute to idealism at the level of history wherein events are represented as "caused by the ideas of men" (Pannekoek, 1937). For traditional cognitive science, as for, Heine, "thought preceded action as lightning did thunder" (Heine, 1832, cited in McLellan, 1969, p.8).

In contrast with this a Marxian account does not maintain that thought has temporal priority over action. The Marxian conception of 'praxis, as we have argued in chapter 2, is one wherein the direction of causation is as often from action to ideas as vice versa, or is one wherein thought and action are held to form a unity. A Marxian approach, we want to say, is precisely not one that emphasizes, in a one-sided way, "working solutions to problems out in the intellect or in the head prior to putting them into practice in the 'real world'" (Torrance, p.53), but rather is one which stresses, in various

ways, that practice itself can provide solutions to theoretical problems. Its outlook thus does not conform to the linear 'sense-model-plan-act' approach of traditional cognitive science.

Planning - GOFAI Marx vs. Praxis Marx

In arguing that a Marxian account is one that does not give temporal priority to thought but instead places emphasis on 'praxis', we are also arguing for a version of Marx which does not conceive of the agent as first and foremost a 'planner'. Our motivations here are partly 'political'. Like Levine we wish to emphasize the "humanistic, creative, spontaneous aspects of Marx" (Levine, 1975, p.xvi), because we believe that the conception of the agent it favours is one that is more conducive to a genuinely free society. However it should be stressed that this is only an interpretation of Marx. We have already seen (in the previous chapter) how Engels had a more planning-oriented conception of the agent, from which he deduced the appropriateness of social planning. Likewise Levins & Lewontin argue that according to 'classical Marxism':

> What seems to be unique to humans is the conscious planning, the imagining of the result before it is brought into existence by deliberate teleological action. (Levins & Lewontin, 1985, p.255)

This interpretation does find potential support in a few passages of Marx's own work, a fact which, perhaps surprisingly, has also been noticed within cognitive science. Margaret Boden (2006) talks of an aspect of Marx's work which she thinks wholly amenable to the project of traditional cognitive science. In fact she goes as far as to assert that "The NewFAI researchers, one might say, had been committed Marxists" (p.1029). Here, by way of clarification, we should point out that Boden prefers to use the term 'NewFAI' – New Fangled AI – to refer to classical artificial intelligence research, as she thinks the term 'GOFAI' has pejorative overtones.[6] The explanation for Boden's attribution of Marxist ideas to classical AI researchers is her belief that Marx too had "stressed the role of anticipatory planning in intelligent labour" (ibid). In

---

[6] Rather confusingly Haugeland (1996, pp.21-28) uses the same term to describe connectionism and embodied/embedded AI.

support of this contention she cites a well known passage from *Capital* Volume 1. We will reproduce this citation, as she has presented it, in full:

> [Man] sets in motion the natural forces which belong to his own body, his arms, legs, head and hands, in order to appropriate the materials of nature in a form adapted to his own need…[So do animals – but human labour is different.] A spider constructs operations which resemble those of the weaver, and a bee would put many a human architect to shame by the construction of its honeycomb cells. But what distinguishes the worst architect from the best of bees is that the architect builds the cell in his mind before he constructs it in wax…Man not only effects a change of form in the materials of nature, he also realizes his own purpose in those materials.   (Marx, 1867, pp.283-4; cited in Boden, 2006, p.1029).

This passage, then, appears to reveal a 'planning' Marx, a Marx who envisions offline cognition as a necessary prelude to online activity. There are possibly some points that one could make by way of mitigation here. One could certainly argue, for instance, that Marxian praxis – in all its variant forms – does not exclude an element of planning per se. Indeed it would be strange if it did for one could then only account for the effectiveness of an agent's actions by recourse to good luck or 'precognition'. The point, we might say, is that planning and action are mixed up in a reciprocal way, with each facilitating the development of each other. As Suchman puts it: "plans are themselves located in the larger context of some ongoing practical activity" (Suchman, 1987, p.49). On this version of things human beings are not first and foremost detached planners, for it no more makes sense to explain action, in one-sided fashion, exclusively as the outcome of planning than it does to explain planning exclusively as the outcome of action.

Our line of defence here, then, is that Marx's apparent emphasis on planning in the above passage is not incompatible with his more fluid 'praxis'-oriented view of the agent. A similar tack also seems to have been adopted by Ollman who suggests that "in Marx's defense it must be added that the terms 'purpose' and 'plan' ('design') are generally used in their weak senses" (Ollman, 1971, p.111), although rather problematically Ollman then goes on to specify that these weak senses include "simply that man knows what it is he is going to produce, and which actions and implements will produce it" (ibid).

However, it does seem that Boden has a point. Insofar as it is possible to derive any account of individual agency from this particular passage, it would appear to be one that supports in a general way a GOFAI perspective – for the notion that "the architect builds the cell in his mind before he constructs it in wax", even if it constitutes an example of Marx exaggerating for effect, seems fairly unambiguous. There is no mention, for example, of the architect formulating the plan in a dynamic way using the tools at his disposal. The planning, as characterized, is as offline, passive and fully comprehensive as it could be and thus presents a rather non-praxis oriented account of agency. We are thus forced into recognizing that just as there are more or less determinist versions of Marx, so there are more or less planning oriented versions of Marx. Here we can only reiterate that, as far as this thesis is concerned, it is the praxis-oriented Marx that is of interest to us.

Alienation from productive activity

We have asserted, then, that the account of the praxis-oriented Marx is at odds with that of traditional cognitive science on the question of the necessary temporal priority of thought, (or more specifically 'planning'), to intelligent action. We have presented this in terms of the Marxian critique of 'idealism'. However it is perhaps also possible to re-describe it in terms of one of the aspects of alienated labour outlined in chapter 2 – alienation from productive activity.

We have already said how, on a Marxian outlook, Man is an "embodied, living, real, sentient objective being" (Marx, 1844a, p.104) who creates himself through his productive engagement with the material world. On this analysis activity is not ancillary to human intelligence but is central to it. We might contrast this with the perspective of traditional cognitive science which tends to view practical activity as mere output and as such not integral to the agent considered as a cognitive being. As Haugeland puts it, "all the 'action' is within the system, rather than being an integral part of a larger interaction with an active body and an active environment" (1996, p.25).

Hence we might say that traditional cognitive science presents an 'alienated' account of an agent's relationship to his activity at least in the sense, as employed by Levins and Lewontin (1985), that it posits an objective *separation* between on the one side 'cognition' and on the other 'activity'. Here we have to be clear about what we are saying. Traditional cognitive science does not describe a scenario wherein an agent is alienated from his activity because such activity has been imposed upon the agent and has certain characteristics inimical to his self-development. What it does do is provide a fragmented conception of the agent which mirrors schematically, and presents as intrinsic to human agency, what a Marxian outlook describes as a temporary product of a particular form of social organisation.

As outlined earlier, this severance of thought from activity – as well as from other aspects of material existence - can also be viewed in relation to the separation of 'head and hand' that arises with the mental/manual division of labour. We might add that it perhaps provides some justification for such a separation. Whilst we do not want to overly concern ourselves with normative matters in this thesis, we might suggest that whereas the Marxian account views man first and foremost as an active producer, and so is equipped to diagnose the cognitive repercussions of particular modes of organizing his productive activity, the traditional cognitive scientific account sees nothing essential in activity or production itself, and so is compatible with an organization of human productive activity which shares this lack of concern. We will return to this topic in the next chapter.

Alienation of Man from Man

If the separation between cognition and action can be aligned to the alienation from activity, this raises the question of whether there are analogues for other aspects of alienated labour in the assumptions made by traditional cognitive science. Whilst, because of its fundamental ambiguity, it would prove unrewarding to attempt this exercise in any depth with 'alienation from species being', there are comparisons to be made with the two remaining aspects of alienated labour. We might first turn to the 'alienation of man from man'.

In fact, any reference to 'social' matters – of inter-agent interaction – was noticeable by its absence from our summary of traditional cognitive science. From this it might be possible to conclude that traditional cognitive science is agnostic on such issues. However it is also possible to see this very absence as significant. As we have said Marx's 'Homo Faber', was not only a productive animal but an essentially *socially* productive animal whose relationship to his own self "first becomes objective to him through his relationship to other men" (Marx, 1844a, p.84). The individual agent, for Marx, is constituted as a social being. For traditional cognitive science, however, the agent is essentially a "thinker". Moreover this thinking is construed as an entirely individual activity, the internal processing of symbols. It may be that the environmental effects of other agents' actions provide inputs to this symbol processing activity, but there is nothing which qualitatively differentiates the way such inputs are handled from the way other inputs are handled, and they have no bearing on the agent's integrity as a self-contained entity. As Froese and Di Paolo put it: "all social phenomena, including social cognition, is reducible without remainder to individual mechanism" (Froese and Di Paolo, 2004, p.3). Speaking loosely, then, we might say that there is no essential connection between one agent and another on a traditional cognitive scientific approach. Just as there is a principled separation between an agent and his activity, so there is a principled separation between agents. We might, therefore, as before, see in this separation a mirroring of the alienated relations posited in Marx's account of alienated labour, in this instance the 'alienation of man from man.'

In some respects, however, the relationship between Marx's alienation of man from man and traditional cognitive science's separation of agent from agent might be viewed as a stronger one than this. Whilst our short summary of traditional cognitive science contained no mention of interactions between agents, if cognitive science is construed in a broader sense as including, or at least having an interest in, developmental psychology, then attempts to describe such relations abound. Insofar as such attempts reproduce the classical assumption of the intrinsic separation of agents, they seem 'alienated' in a more palpable sense. Thus, for example, Harnad states that:

> It has been found that children after a certain age, and certain animals, have considerable skill in detecting or inferring what others (usually members of their own species) are feeling and thinking… Let us note right away that this sort of mind reading is a form of Turing-testing: inferring mental states from behaviour…We know… since

at least Descartes, that the only mind that we can read other than by Turing-testing is our own. (Harnad, 2003, pp.72-3)

The assumption made here is similar to that made in philosophical accounts of the 'problem of other minds' and in a proposed solution to that problem known as the 'argument from analogy' (see e.g. Mill 1889). It is also more directly a reference to the perspective adopted in psychological accounts of 'theory of mind' (see e.g. Baron Cohen et al, 2000), and in particular a version of (or explanation) of theory of mind known as 'theory theory'. In all of these accounts the unquestioned starting point is that of a problematic separation between individual agents which goes beyond mere physical separation. The suggestion is that whereas agents have direct access to their own mental states they have no 'direct' access to the mental states of others. A solution is then posed in terms of detached inferential mechanisms which take outward behaviour as evidence for inner mental states.

Such accounts, we might say, bring to mind the subjective aspect of Marx's 'alienation of man from man', a point that some commentators have already made in relation to philosophical 'other minds scepticism' (e.g. Fann, 2002). The assumption of a fundamental separation between individuals makes the normally unreflective interchanges between those individuals seem strange and problematic – this is implicit in the very term 'mind-reading' which invites (if only by suggestion) comparison between the predicament of the average human being when trying to understand his conspecifics and that of the telepath who uses his powers to gain direct access to another's mind. On these accounts what is, on a Marxian analysis, man's "essential nature" (Marx, 1844b, p.122) as a social being, appears as something mysterious and in need of explanation via recourse to an external mechanism.

A Mechanical Robinson Crusoe

The principled separation of the agent from other agents, then, mirrors the alienation of man from man, and when applied in the field of developmental psychology gives rise to analyses which appear very much like descriptions of the subjective alienation of man from man. We might also note that this separation can be seen as generating a further issue for traditional cognitive science, and one which has some relevance to a Marxian

outlook. This issue will not be pursued in depth as to do so would involve opening a whole can of worms on the subject of the social character of 'representation'. Nevertheless it is worth giving a brief summary here.

Marx's perspective, as we have seen, is one which does not take the "isolated individual" (Grundrisse, p.83) as its starting point, for the individual, considered as anything more than a physical organism, is a social entity – "the human being is … not merely a gregarious animal, but is an animal which can individuate itself only in the midst of society." (ibid). In this context the attempt, by traditional cognitive science, to pitch explanations of cognitive phenomena only at the level of the mental processes of the individual agent can be viewed as being as flawed as the attempts by classical economists to project economic qualities onto the isolated individual. Adam Smith's conception of 'Homo Economicus' , for example, was that of an individual who confronted the world as a fully formed free marketeer with a natural "propensity to truck, barter and exchange" (Smith, 1776, p.117). According to Marx such accounts belonged "among the unimaginative conceits of the Robinsonades" (Marx, 1855, p.83) i.e. they depend upon a 'Robinson-Crusoe-like' conception of the individual, inexplicably fully formed with social properties yet without recourse to social bonds.

We have, then, a parallel between individualized economic qualities and individualized cognitive properties. We might note that in both cases reference to a symbolic medium is implied. With Smith's 'Homo Economicus' his predisposition to exchange is dependent on exchange value as manifest ultimately in a social, symbolic medium of exchange – money. With the GOFAI agent, cognitive abilities reside in formal symbol manipulation. We might add to this mix Marx's reference to a third representational medium – language. In considering the possibility of production performed by a single individual, living outside of society, Marx came to the conclusion that this would be:

> …a rare exception which may occur when a civilized person in whom the social forces are already dynamically present is cast by accident into the wilderness – [but] is as much an absurdity as is the development of a language without individuals living together and talking together.   (Marx, 1855, p.84)

The comparison of individualized production with individualised language, and the judgement that both are absurd, anticipates to some degree Wittgenstein's private

language argument. In this respect we might note that Wittgenstein also drew parallels between the idea of individualized economic properties and that of individualized linguistic abilities. Thus, for example, he attempts to demonstrate the incoherence of the notion of a necessarily private language by asking "Why can't my right hand give my left hand money?" (Wittgenstein, 1953, p.94) In response to this, it is perhaps significant that individualist oriented defences of a private language again invoked 'Robinon Crusoe' scenarios. Thus, for example, Ayer:

> Imagine a Robinson Crusoe left alone on his island while still an infant…He will certainly be able to recognize many things upon the island…Is it inconceivable that he should also name them?...Surely it is not inconceivable that someone uninstructed in the use of an existing language makes up a language for himself. After all, some human being must have been the first to use a symbol. (Ayer, 1954, p.259)

These threads might be brought together in the following way: Just as traditional cognitive science thought cognition explicable in terms of the internal symbol processing powers of the isolated agent, so classical economists (or at least some of them) saw an individual 'innateness' in the propensity of the agent to manipulate external symbols (money);[7] and likewise again so pro-private language theorists saw no contradiction in the idea of linguistic symbol manipulation in a vacuum. In each of these instances the essential practical-social character of symbol usage was ignored.

In the case of GOFAI this manifested itself as the "symbol grounding problem" (Harnad, 1990) – the problem of how "the meanings of … meaningless symbol tokens, manipulated solely on the basis of their (arbitrary) shapes [can] be grounded in anything but other meaningless symbols." (p.335.) The "pure thought" of idealist philosophy which, on a Marxian analysis, exists in complete detachment from the realm of 'sensuous activity' and so has no chance of forming an adequate conception of its object, thus finds an earthly counterpart in the meaningless tokens of GOFAI. Like a mechanical Robinson Crusoe the agent as envisioned by traditional cognitive science is doomed to processing in a symbolic medium that has no purchase on the real world.

---

[7] Also of relevance here is Marx's remark that "Logic is the money of the mind" (Marx, 1844a, p.99).

Alienation from the Product

We have argued that there are analogues for the alienation of man from his activity and for the alienation of man from his fellow men, in traditional cognitive science. We might pursue this line of enquiry by asking if there is anything in traditional cognitive science corresponding to the alienation of man from the product of his activity.

In fact our earlier suggestion that the separation of cognition from activity - where such activity is considered as mere 'output' – 'mirrors' the alienation of man from his activity, perhaps already implies that a similar story can be told about the 'product' of this activity. Indeed, insofar as 'product' and 'output' are synonyms the exact same story can be told. Whatever takes place after cognition, or in separation from cognition, whether or not it materially alters features of the external world, is peripheral to cognition on the traditional account. Hence the separation of cognition from output could also be said to 'mirror' the alienation of man from the product of his activity.

However there is a more obvious and substantial equivalent of the 'alienation of man from his product' to be found in traditional cognitive science - the notion, outlined earlier, of 'Strong AI'. The suggestion that appropriately programmed computers literally have minds, whatever exactly that might mean, seems to reproduce the essential features of the alienation of man from his product. This is so, we might suggest, not only in the restricted sense – implied in Marx's passages on alienated labour – where the product in question is the immediate result of the agent's activity, but also in the broader sense which underlies both Hegel's and Marx's various accounts of the alienation of the subject from his object. In all of those accounts man (or consciousness) externalized him/itself in his/its products – which might equally be God, or the commodity as a recently completed artifact – and then confronted those products as something 'other'. Likewise we might say that in the case of 'mind-like' artifacts it is the intelligence of the agent which is transferred to the artifact and becomes *its* intelligence. As with those other examples of externalization, the nature of this transference remains hidden from the agent, who views this human product as endowed with a life and energy of its own.

We might also draw perhaps comparisons specifically with Marx's account of commodity fetishism at least insofar as, in that scenario, the alienation of man from man was accompanied by the projection of human properties onto objects. In the case of traditional cognitive science, as we argued earlier, an individualist conception of the agent makes intersubjective relations problematic; nevertheless it is at the same time willing to consider the application of mental predicates to artifacts. Thus, albeit speaking imprecisely, we might say that in traditional cognitive science an 'other minds scepticism' about human agents is accompanied by an 'other minds liberalism' about machines.

Viewing 'Strong AI' as an instance of the alienation of man from his product perhaps puts a new slant on John Searle's arguments. Searle attempted to demonstrate that "such intentionality as computers appear to have is solely in the minds of those who program them and those who use them" (Searle, 1980, p.199). Like Hegel, we might say, Searle believes that "behind the curtain of appearance is not an unknown thing-in-itself but the knowing subject" (Hegel, 1807, p.112). The mind of the computer is our own mind.

However there are respects in which Searle's argument is perhaps not entirely harmonious with a Marxian perspective. Although he does at one point come quite close to putting his analysis of Strong AI in a wider context of externalization through labour, noting that the general practice of attribution of intentionality to artifacts "has to do with the fact that in artifacts we extend our own intentionality; our tools are extensions of our purposes.." (Searle, 1980, p.188), this analysis is never developed. Instead he prefers to view such attributions as mere "metaphor and analogy" (ibid. p.180) on the part of the layman, with the implication that the Strong AI enthusiast has made the simple mistake of taking such figurative speaking literally.

We might contrast Searle's approach here with that of Dennett. His 'intentional stance' view, on one reading, comes perilously close to Strong AI. For Dennett any organism or artifact which can be usefully treated as a "rational agent" *is*, for all intents and purposes, a rational agent. Thus: "Any system whose behaviour is well predicted by this strategy is in the fullest sense of the word a believer" Dennett, 1981, p.59). On this

basis Dennett argues that different degrees of intentionality might be rightfully attributed to everything from thermostats and clams up to sophisticated robots and human beings. However this is not mere detached rationalistic prescription on Dennett's part – for he sees himself as describing a pre-existing human practice, the justification for which lies in the fact that it works: "Even when we are surest that the strategy works for the wrong reasons, it's nevertheless true that it does work, at least a little bit" (ibid p.66). To this extent Dennett's position also has something in common with a Marxian account insofar as it likewise roots such attribution in human social practice and material circumstance. It is, we might say, as difficult to avoid adopting the 'intentional stance' as it is to avoid externalizing oneself in one's products.

Thus a Marxian view can perhaps be seen as incorporating features of both accounts. Searle's position is correct, it would be argued, insofar as it reveals the intentionality of the artifact to be nothing more than the intentionality of the (human) agent, unrecognized as such; but it is wrong insofar as it portrays such projection as mere accident or habit. Dennett's view, by contrast, is correct insofar as it suggests that there is something materially significant in the compulsion to attribute different degrees of intentionality to artifacts of all sorts, thus illustrating that the 'Strong AI' hypothesis is not different in kind to the general practice of applying a more restricted set of mental attributes to thermostats etc. However it is wrong insofar as it implies the literal truth of artifactual intentionality. From the standpoint of alienation the compulsion to attribute intentionality to a range of artifacts is explained (in part) by experience of such artifacts as separate powers. As we saw in chapter 2, insofar as an artifact is an externalized product of another human being or group of human beings, operates by means of a mechanism that is not understood by the 'user' (or worker) and imposes the logic of that mechanism on the lived experience of that user, it is viewed as an objective intelligence. This is what Wendling referred to as "machine fetishism" (Wendling, 2009, pp.55-60). This tendency, we might argue, takes on its most concrete, extreme and literal form in the case of Strong AI where human intelligence manifests itself as a complete, self contained 'other' and is consciously feted as such by human agents.

In putting forward such a perspective, however, we do not wish to imply that Dennett's account of the intentional strategy should be replaced by an analysis which emphasizes only human alienation from 'intentional' artifacts. It is difficult to quibble with that

portion of Dennett's account which seeks to show the utility of the application of intentional terms to artifacts. Moreover Dennett is free to argue that this usage of intentional terms is as legitimate as any other – that it is not simple 'metaphor', as Searle has suggested. However his suggestion that these terms can be applied to artifacts *in the exact same sense* as they are applied to human beings is what is in dispute, or from a Marxian standpoint we might say that it is only correct in the convoluted sense that such artifacts are the products of externalized human intelligence. That man externalizes himself in his artifacts does not imply that those artifacts literally have mental properties, a point we shall return to in Chapter 7.

Mechanism

We have suggested various senses in which traditional cognitive science might be characterised as 'alienated' or as offering an alienated conception of the agent. These have included its general idealist orientation - in particular its emphasis on the temporal precedence of ideas over action - and its mirroring to different degrees the alienation of the agent from his activity, the product of his activity and his fellow agents.

However, although some of these features can be related to a 'computational', and hence loosely 'mechanistic' account of mind, and whilst we have already described the GOFAI agent, in passing, as a 'mechanical Robinson Crusoe', we have not yet broached the topic of 'mechanism' per se.

As we have seen, both Hegelian and Marxian accounts tended to be pejorative about "mechanistic" perspectives in general. It is true that it is not always clear that there is unanimity in the definition of 'mechanism' underlying such critiques. Some Marxist critics of mechanism - particularly Engels - have themselves been accused of a 'mechanistic' outlook by other Marxists (e.g. Levine, 1975) suggesting a degree of fluidity in the use of this term. Nevertheless, given that traditional cognitive science has as its explicit project the representation of 'mind as machine' (Boden, 2006) we need to ask if there are any clearly specifiable Hegelian/Marxian senses in which a critique of

traditional cognitive science might be related directly to a critique of its 'mechanistic' nature.

The Cognitive Production Line

One relevant feature of traditional cognitive science in this respect is its commitment to a linear, unidirectional conception of cognitive processing. This was evident, for example, in Brooks' (1991a) 'sense-model-plan-act' cycle. Van Gelder provides a succinct summary of cognitive systems viewed from a similar standpoint:

> They interact with their environments in a cyclic process that begins with input transducers producing symbolic representations in response to the environment, continues with sequential internal computations over symbolic structures and ends with output transducers affecting the environment… (van Gelder, 1996, p.438)

What can perhaps be argued here is that such architectures reflect, to a certain degree, the organisation of tasks in the 'detailed division of labour'. Like a factory, we might say, the cognitive system has its input of raw materials and outputs a finished product. In between are various stages of processing which may be envisaged on a par with alterations made to the developing product on a production line. Of course such a conception of cognitive processing has its most obvious roots in the serial processing of a digital computer rather than in the detailed division of labour. However, as we saw in chapter 2, the two are not unrelated for both were developed in tandem in the work of Charles Babbage. Thus, whilst the connection between the detailed division of labour and a serial processing conception of cognition might not be a straightforward instance of a mental model "mirror[ing] the structure of the alienated social world in which it was conceived" (Levins & Lewontin, 1985, p.270), there is a significant material linkage here. The isomorphism is not coincidental.

Minsky

It should be noted at this juncture that the exercise of critically relating certain conceptions of mind to particular modes of organising production, is not something new to cognitive science. Marvin Minsky (1985), for example, put forward an explicit critique of a 'factory' oriented conception of mind:

> In earlier times, we could usually judge machines and processes by how they transformed raw materials into finished products. But it makes no sense to speak of brains as though they manufacture thoughts the way factories make cars. The difference is that brains use processes that change themselves – and this means we cannot separate such processes from the products they produce. (Minsky, 1985, p.288)

Here, we might say, Minksy occupies an interesting middle ground between 'alienated' and 'unalienated' conceptions of mind. His desire to highlight and debunk a factory metaphor of mind shows some affinities with a Marxian outlook. Moreover his suggestion that 'brains use processes that change themselves' is in keeping with the Marxian notion of the self-creation of the agent through reciprocal interaction with the environment. It also, we should note, prefigures some 'enactivist' ideas (see chapter 6) and indeed, Varela, Thompson and Rosch make extensive use of Minsky in outlining their own position (e.g. V.T.R., 1991, pp.138-9). However, the sticking point is perhaps Minsky's focus on the brain itself. It is the brain and not the agent which constitutes itself through its own productive processes. Hence insofar as Minsky sees the brain as the locus of action – in isolation from embodied, social, "sensuous human activity" (Marx, 1970a, p.121) – his analysis can still be said to reinforce the fragmentary outlook of traditional cognitive science.

Dahlbom

Minsky is not the only cognitive scientist to examine mind in relation to industrial production. Another is Bo Dahlbom. Dahlbom is more explicit than Minsky in postulating a causal link between modes of social/productive organisation and models of mind, going so far as to suggest that theoretical conceptions of the mind are typically "more revealing about the society of their protagonists than about the human organism" (Dahlbom, 1993, p.3). One of his targets is Dennett's (1991) 'multiple drafts' theory of consciousness:

> 'Multiple drafts' – what a wonderful way to summarize the anarchistic liberalism of a free market! Dennett is inviting us to apply "postfordism" to consciousness, to give up thinking of mind as a centralized, bureaucratic organization of Ford production lines, and begin thinking of it, rather, as a decentralized, flexible, organic organisation. (Dahlbom, 1993, p.4)

The political outlook underpinning Dahlbom's criticism of Dennett, we might note, is different to that underpinning this thesis. For Dahlbom there appear to be only two

possible social models – a planned economy or a free market economy (and he seems to favour some version of the former.) Hence accounts of mind which dispense with centralization are thought to be allied to a competitive free market outlook. Whilst this might be true of Dennett's own outlook, it ignores the existence of a third possibility – that "decentralized, flexible, organic organisation" might be compatible with the absence of a market altogether. It is such a reading of Marx – one which emphasizes a free society without centralized planning, nor any kind of market – which informs this thesis.

Politics aside however, given that we accept, in accordance with a Marxian outlook and in agreement with Dahlbom, that modes of social/productive organisation are reflected in models of mind – what does this tell us about the viability or otherwise of those models of mind? In our own case, if a serial processing conception of cognition can be convincingly linked to the detailed division of labour, does this count for or against such a conception (or neither)? Dahlbom's own answer to this question is rather surprising. Starting from the general premise that minds are embedded in societies he suggests that "if mind is a social phenomenon rather than a brain process, then the use of social concepts in a theory of mind may not be metaphorical after all" (ibid. p.6). He fleshes this position out as follows:

> Theories like … Dennett's Multiple Drafts model are more obvious expressions of contemporary American self-understanding: free competition in a free market. To the extent that they are true of that society, it is likely that they are true of the minds in that society…If mind is in society there is no difference here.   (ibid)

This argument seems to present some difficulties. Firstly there is the problem that, although members of a particular type of society might be more predisposed to project particular organizational assumptions onto their mental models, there is no reason to think that these mental models might be ones which have any practical connection or compatibility with life in that society. Citizens of a very hierarchical society might be inclined to think that mind is also hierarchical, but a hierarchical mind and a hierarchical society may have nothing in common apart from a certain formal isomorphism at a particular level of description. Thus, for example, even if "mind is social" there are no 'a priori' grounds for thinking that "hierarchical minds" are best

suited for life in a hierarchical society. (We are of course being deliberately vague about what sort of thing a "hierarchical mind" might be.)

Secondly, and connectedly, there is the danger that Dahlbom's argument conflates "having a particular conception of mind" with "having a particular type of mind". We are in agreement that certain modes of social organization are more likely to give rise to certain conceptions of mind but having a particular conception of mind need not be having a certain type of mind. In atomized societies, for example, we may be more predisposed to think we have (or could have) a 'private language', but this is not the same as actually having a private language. That we think we have a private language could be mere 'false consciousness' from a Marxian standpoint. (This is not to say that there might not be certain mental states - e.g. anxiety - where thinking that you have it is the same as having it.)

Lastly, and again connectedly, there are at least some features of 'mind' which, even given facts about our susceptibility to environmental influences, are not likely to be dependent on social organization or social agreement. Neither, however, need they be adequately characterized as "brain processes" (ibid). One such feature might be cognitive architecture. It seems plausible to suggest that if cognition *were* internal symbol processing then no re-organization of society or productive relations would be likely to change this fact. At the same time, although symbol processing might be realized in a brain, it cannot be *identified* with a brain process, at least insofar as, according to the principle of multiple realizability, it might equally be realized elsewhere, or in a different way.

Dahlbom's argument, then, does not give us any grounds for supposing that a 'production line' conception of cognition could become true if it were not true in the first place. This, however, does not bring us any closer to understanding whether, from a Marxian perspective, such a conception *is* viable. We have so far only said that such a view mirrors structurally a certain mode of production. We have yet to specify what it is about its mechanistic nature which warrants Marxian disapprobation.

Reciprocity vs uni-directionality

Answering this question involves returning to ground already covered in earlier discussions of planning and praxis. We said that, from a Marxian viewpoint, traditional cognitive science was susceptible to some of the same criticisms as idealism, because it likewise gave priority to thought. This 'priority' might be considered in two ways. One is just the general sense in which thought is held to be the most significant feature of an intelligent agent's being; the other is the more specific sense in which thought (cognition) is held to be temporally prior to action. In countering this latter conception we said that a praxis oriented view was one that did not give thought necessary temporal priority over action. Action is as much a condition for thought as vice versa, and in this way it might be said that action and thought form a reciprocal 'unity'.

If we turn now to a serial processing view of cognition, considered from the standpoint of mechanism, the same point might be made again but with less reference to 'thought' and 'action', and with a greater emphasis placed on countering uni-directionality per se. We have seen in chapter 3 how both Engels and Merleau-Ponty presented criticisms of a mechanistic outlook. Engels sought to differentiate "mechanical materialism" (Engels, 1887, p.321) from a dialectical version of materialism on the grounds that the latter incorporates a reciprocal relation between qualitative and quantitative change. Merleau-Ponty similarly distinguished his own dialectical account of agent/environment interaction from one based on "mechanical action" (Merleau-Ponty, 1942, p.160) on the grounds that the latter could not accommodate genuine reciprocal interchange between agent and environment. On both accounts, then, albeit in slightly different ways, 'mechanism' is partly characterized in terms of uni-directionality and lack of reciprocal causal relations.

On this basis we might say that one of the ways in which serial processing views of cognition, as exemplified in the 'sense-model-plan-act' architecture, are mechanistic in a negative Marxian sense, is that in which they too are unidirectional and fail to account for a more fluid reciprocal interaction between agent and environment. Although such architectures are cyclical insofar as actions give rise to updated sensory input which can

then be used to generate plans for additional actions (Russell & Norvig, 1995, pp. 31-52), this fragmented cyclicity cannot be identified with genuine reciprocal interaction because at each point in the proceedings we can isolate discrete episodes which, in uni-directional fashion, are the "necessary and sufficient condition" (Merleau-Ponty, 1942, p.160) for the subsequent episode in the sequence.

Discreteness and continuity

In arguing that a serial processing view of cognition is mechanistic, and therefore undialectical, because it does not accommodate reciprocal causal relations, we have made reference to the discrete episodes that compose it. Although the dichotomy between discreteness and continuity is integrally bound up with that between linear causality and reciprocal causality, we might treat 'discreteness' here separately, as a second component of a mechanistic outlook attributable to traditional cognitive science.

Discreteness is at the heart of traditional cognitive science. The "discrete symbols" (Eliasmith, 2001) of a physical symbol system require the operations of a "discrete state machine" (Turing, 1950), such as a computing device, to process them. Such a machine is one that:

> [moves] by sudden jumps or clicks, from one quite definite state to another. These states are sufficiently different for the possibility of confusion between them to be ignored. (Turing, 1950, p.35)

We might notice two things about this description. Firstly, the character of the transitions of the discrete state machine is not held to be important to the machine or to the user of the machine. It is not informationally significant. All that one needs to know is that a transition has been made. That they are 'sudden jumps or click' emphasises their status as empty spaces between states. Secondly, these states themselves stand in separation from each other and are independently specifiable as such.

A dialectical critique of such an account - as a description of the workings of the human mind - is partly suggested by Turing's own later remark that "strictly speaking there are no such machines. Everything moves continuously" (ibid. pp.35-6). This comment,

possibly deliberately, echoes the maxim attributed to Heraclitus that "everything is in flux" (Barnes, 1987, p.114-117),[8] a maxim taken up as a guiding principle by both Hegel and Engels. Their dialectical analysis, as we have seen, was one which understood the world as in "constant motion, change, transformation, development" (Engels, 1887, p.37).

However there is more to a dialectical critique of discrete state transitions than the mere acknowledgement that things move and change in a continuous fashion. In this connection we might recall the progressions described in Hegel's *Logic*. These can be regarded as a series of state transitions, at least in the sense that 'states' of knowledge are implied at each juncture. One thing we note about such states is that they are not discrete. They cannot be understood in separation from each other but rather each state brings forth the next and the latter state is only what it is in relation to the former. In this way, 'Being' and 'nothing', for example, are internally related.

Moreover, not only are these states not discrete but the transition from one to the other is held to be as significant as the states themselves. Thus the transition to a third state, 'becoming', only arises when mind considers the transition from 'being' to 'nothing' and sees in it a 'becoming'. Here, the importance of 'transition' itself is underlined by the fact that the earlier transition from 'being' to 'nothing' is taken into account as significant and a determining factor of the successive state. In this particular case it is also underlined by the fact that the successive state is itself labelled 'becoming', thereby expressing the centrality of 'transition' to Hegel's dialectical outlook. It is the transition "which is the essential point", as Hanna puts it (Hanna, 1996, p.228). Somewhat figuratively we might add that if 'being' and 'nothing' correspond to the ones and zeroes of a digital machine, 'becoming' emphasizes the importance of the transition from the one to the other.

We might take it, then, that a dialectical critique of mechanism as applied to cognition implies also a critique of the discrete state transitions upon which symbol processing depends. A dialectical outlook emphasizes not only continuity and change in a general way, but also the dependent relations between states ('determinations') and the central

---

[8] A version of this maxim is attributed to Heraclitus by both Hegel and Engels, but is not to be found in surviving fragments of his work.

importance of the transitions themselves. The discrete state transition perspective could therefore be characterized as 'alienated' at least in the sense employed by Levins and Lewontin (1985, p.270) that it presupposes the essential separation of the relevant states, and so imposes an unwarranted fragmentation on cognition.

Bivalence and Holism

In our analysis of 'mechanism' in traditional cognitive science we have so far looked at two features of a 'production line conception' of cognition – unidirectionality and discrete state transitions – and have said that both are at odds with a dialectical approach. This does not, however, exhaust the ways in which traditional cognitive science might be viewed as mechanistic in a sense that an Hegelian or Marxian perspective could find objectionable. We might remember Hegel's critique of the "judgements and syllogisms" of formal logic, that the rational operations they facilitated were "completely analytical…and mechanical" (*Science of Logic*, p.52 cited in Hanna, 1990, pp.269-270). Implicit in this critique were two arguments that are at least separable in theory, a critique of the bivalence of standard logic and a critique of a lack of holism in a "language of symbols" (ibid p.270).

As regards the former, insofar as computers can be viewed as mechanized formal systems (Haugeland, 1996, p.11), incorporating in particular the principles of standard bivalent logic, and insofar as traditional cognitive science involves the positing of this mechanization as a means by which the essentials of human cognition can be understood, it would seem fair to suggest that the latter might be offering a conception of Mind or Reason antithetical to that of Hegel. GOFAI, we might say, has taken the non-dialectical 'either/or' of the Understanding and has used it to create artificial minds which are then taken to be reflections of our selves.

As regards holism, we have already said how a notion of continuity in, and internal relatedness between, states or components of a process, forms part of a dialectical critique of mechanism. These are all implicated in the 'temporal holism', outlined in chapter 1, where, on an Hegelian account, to understand something in its totality is to understand it both in relation to the whole of which it is currently a part, and as a

process over time (Rees, 1998, p.46). With this viewpoint in mind we might say that an Hegelian critique of Newell's 'physical symbol system hypothesis' would be the same as that which he directs at Leibniz's "calculus of combinations and permutations" (Hanna, 1990, p.270). A physical symbol system is unable to provide an adequate representation of the world because it abstracts from that world certain features which are then reconstituted in fixed and isolated fashion as symbols. The mechanical re-animation of a collection of such symbols cannot capture the world considered as a "realm of organic relationships, dynamic processes and concrete truth" (ibid) because those symbols lack the requisite fluidity and connectedness.

Base or Superstructure?

We have attempted, then, to give some indication of senses in which traditional cognitive science might be considered mechanistic in a way that runs counter to the outlook of Hegelian-Marxian theory. We have said that insofar as it has a conception of cognition that is unidirectional, is founded on the presuppositions of bivalent logic and involves discrete operations on fixed symbols, it is intrinsically undialectical in nature. Where dialectics emphasizes reciprocity, continuity and holism, GOFAI emphasizes linearity, fixity and discreteness.

It should perhaps be noted, however, that we have only touched upon these issues and that further probing might reveal a wealth of unanswered questions. It is easy, for example, to talk in a general way about the mechanization of standard logic forming the groundwork upon which the edifice of traditional cognitive science is erected, but things get a little trickier if we try to be specific. If, for instance, our concern is with the inherently bivalent nature of standard formal logic, are we referring to the incorporation of bivalent logical principles in the very structure of a computing device (e.g. in its use of 'And gates' and 'Or gates'), to the incorporation of these principles into the programming languages used by such machines (e.g. the "if, else" structure in high level languages) or to analogues of such structures and languages in the human mind or brain?

Moreover, we need to make a clear distinction between the bivalence and discreteness at the heart of mechanical computing, and the implementation of programs which may appear to lack these features. It seems arguable, for example, that there are 'dialectical' elements, elements which transcend the 'one sided' perspective of standard formal reasoning, to be found in the holism of connectionist networks (Rumelhart, McClelland, et al., 1986), in the triadic evolutionary strategy of genetic algorithms (Holland, 1975) and in instances of artificial 'emergence' in cellular automata (Dennett, 2003). Each of these can be implemented on standard computer hardware using standard programming languages.[9]

There is nothing prima facie problematic about either of these instances. However it should be noted that there is often a fundamental ambiguity in cognitive science concerning the level of explanation at which any particular model of cognition is pitched (Fodor & Pylyshyn, 1988). This ambiguity is further exacerbated by the fact, as stated earlier, that such models are used to describe both machines and minds, and it is not always clear how one transposes from the one to the other. Insofar as this ambiguity is intrinsic to philosophy of cognitive science it probably also infects our own analysis above.

Conclusion

We have said that traditional cognitive science has something in common with the Marxian version of Idealism insofar as it places emphasis on the agent as a thinker and gives thought temporal priority over action. We have said too that it can be characterized as alienated insofar as it reproduces elements found in Marx's account of alienated labour. Lastly, we have said that its 'production line' conception of cognitive processing not only mirrors the detailed division of labour structurally, but is intrinsically mechanistic, and so undialectical, in several respects. In the next (and subsequent) chapters we will consider more recent cognitive scientific accounts which,

---

[9] We might also note that the author of this thesis once experimented with the implementation of a 'dialectical engine' – an attempt to mechanize non bivalent dialectical reasoning – using the same resources – Queen Mary, 2001.)

it will be maintained, offer a less alienated account of the agent to the extent that they avoid various of the above pitfalls.

Chapter 5 – The Retreat from Alienation

A spectre is haunting the laboratories of cognitive science, the spectre of embodied cognition.

(Goldman & De Vignemont, 2009)

Introduction

Over the last 25 years or so there has been a well documented (see e.g. Clark, 2001; Boden, 2006; Froese, 2007; Robbins & Aydede 2009) sea change in cognitive science. Traditional approaches to mind and cognition have found a rival in a collection of views which stress that mind and cognition are in various ways 'embodied', 'extended', 'embedded' and 'enacted'. It is our central contention that this development might usefully be characterised as a "retreat from alienation."

In characterizing current cognitive science in this way we wish to draw attention to certain parallels between themes and perspectives found in Hegelian/Marxian theory and those found in recent cognitive science. Our intent in doing so is not just to indulge in an exercise in arbitrary and selective pattern matching. Rather we wish to show that viewing recent developments in (philosophy of) cognitive science through a Marxian prism can be theoretically useful. It both allows us to see a common thread in an otherwise disparate cluster of perspectives (see below) and may help us to diagnose weaknesses and limitations in some of those perspectives (see below & chapters 6-8.) As regards this latter point, however, it should be made clear that we are not proposing the inviolability of an all-encompassing and omniscient Marxian perspective - the 'diagnostic' exercise can work both ways, with recent theoretical advances in cognitive science also helping to expose some limitations of a Marxian perspective (see e.g. section on 'Embodiment and Materialism', below) or providing more concrete accounts of ideas that are to some degree ambiguous on Marxian accounts (see e.g. section below on 'van Gelder', and chapter 8 on epistemic action).

As well as proposing the utility of a Marxian perspective on cognitive science we are also proposing a degree of historical justification for such an approach. As has already been suggested in chapter 3, in some cases it can be shown that there is a direct line of

descent from Marxian ideas to certain ideas in current cognitive science, and therefore that the influence of German philosophy on current cognitive science is not restricted to the work of Heidegger (see e.g. Kiverstein & Wheeler, 2012). We will not treat this idea separately in the following, but will allude to it in particular cases where appropriate.

From Separation to Unification

If there was one theme which ran through our account of traditional cognitive science and our diagnosis of it as 'alienated' in a Marxian sense, it was that of 'separation'. Mind, on the traditional account, was viewed as essentially separate from the material world (whether that material world be taken as the body of the agent, or external nature), from action in that material world, and from the social world. In this way it reproduced the separation experienced by the Marxian productive agent – characterized in the *Economic and Philosophic Manuscripts* as alienation from productive activity, from the result of that productive activity and from fellow productive agents (Marx, 1844a, pp 75-111). Similarly, insofar as traditional cognitive science could be said to have something in common with Idealism – in its elevation of cognitive processes above material being, and in its assumption of the temporal priority of thought over action – this too could be said to be expressive of a separation, the separation of thought from "sensuous human activity" (Marx, 1970a, p.121), which on a Marxian analysis has its roots in the separation imposed by the mental/manual division of labour. Lastly, separation was also implicit in a mechanistic ('production line') conception of cognitive processing, which was unidirectional and involved discrete operations on fixed symbols. Such an account was at odds with a dialectical approach which stressed unity of cause and effect ('reciprocity' – Hegel, 1817a, p.217) in interactions between agent and environment, emphasized continuous process in the place of discrete state transitions, and maintained a holistic view of the world resistant to representation by a mechanized "language of symbols" (Hanna, 1990, p.270).

In contrast to this, if there is one theme which unites approaches in the 'new' cognitive science it is supersession of this fragmented outlook. Like Hegel's Spirit, the new cognitive science is against 'separation' and for 'unification' – the unification of the

human subject. Like Marx, modern cognitive science has as its aim the reconstruction of the "whole man" (Marx, 1844a, p.91).

This striving for unity is manifest most obviously in the content of various accounts which portray the human agent (or mind) as embodied, enactive, embedded or extended. Thus we might say that 'embodied' perspectives aim at superseding the dichotomy between mind and body. Contra the idealist orientation of traditional cognitive science, they too declare that "man is an embodied, living, real, sentient, objective being" (Marx, 1844a, p.104). Examples, amongst many others, include Gallagher, whose comprehensive survey of the field makes extensive reference to Merleau-Ponty and has the ultimate aim of showing that "nothing about human experience remains untouched by human embodiment" (Gallagher, 2005, p.247), Lakoff (Lakoff, 1980; Lakoff & Johnson, 1999; Gallese & Lakoff, 2005), who offers a variety of theses relating embodiment to concept formation (see below), and Damasio (1994, 2000) whose account of the role of emotions in rational decision-making utilizes an embodied conception of the former (see below).

Turning from 'embodied' perspectives to 'enactive' ones, we might say that the latter aim, in various ways, at superceding the dichotomy between cognition and action, and so run counter to that alienation of the agent from his productive activity which we found to be mirrored in the outlook of traditional cognitive science. Examples here include Kirsh and Maglio's (1994) concept of 'epistemic action' which can be viewed as undermining the idealist assumption of the temporal priority of thought by positing a praxis-like unity between thought and action (see chapter 8). Another example is Noë's account, influenced by Merleau-Ponty (1945), which argues that "perceiving is a way of acting" (Noë, 2004, p.1) which points to a unity of sorts between perception and action. We will examine Noë's account shortly.

In making the claim that Noë's enactivism adopts a unifying strategy by attempting to supercede the dichotomy between action and cognition we should be aware (as Torrance, 2005, points out) that the term 'enactivism' is often also used to describe a wider perspective which, whilst it incorporates a critique of the action/cognition dichotomy ("All doing is knowing, all knowing is doing", Maturana & Varela, 1987, p.26) is centrally concerned with several other theses. Some of these theses have a

strongly Hegelian/Marxian flavour, and are to some degree descended from the Marxian accounts of Merleau-Ponty and Lewontin. We will discuss this further in chapter 6.

If 'embodied' perspectives aim at unifying mind and body, and 'enactive' perspectives (on a restricted definition) aim at unifying action with cognition, then 'embedded'/ 'extended' perspectives aim at unifying agent and world, whether the latter is conceived of as the world of other agents (Thompson, 2001; Gallagher, 2005; Gallese 2001, 2003; Gallese & Lakoff, 2005) or as the material world (Clark& Chalmers, 1998, 2008; Rupert, 2009). Thus, for instance, Gallese's account (see below) calls into question the natural necessity of the 'alienation of man from man' by positing an active, practical and pre-theoretical linkage between agents, whilst Clark and Chalmers 'extended mind' hypothesis recalls elements of the Hegelian-Marxian concept of externalization by suggesting that "mind extends into the world" (Clark & Chalmers, 1998, p.12).

In some accounts this striving towards unification is consciously articulated as such and takes place on more than one front. Thus Clark (1997), in a wide-ranging account which draws in part on a particular interpretation of Vygotskyan notions of 'scaffolding', has as his explicit purpose "putting brain, body and world together again", a theme which is also clearly presented in some dynamical accounts (e.g. van Gelder, 1995, 1996 – see below). In other accounts it is less well articulated and more restricted in its scope, and perhaps less successful. As regards this latter point we will argue below that Noë's (2004) attempt to unify perception and action is more thoroughgoing and successful – and in some sense more 'Marxian' - than Damasio's (1994) attempt to unify emotion and reason. Nevertheless, whether or not the striving for unity is consciously articulated, and irrespective of the differing degrees of success met by different practitioners, it is our contention that the various projects of many of those involved in the field of philosophy of cognitive science, when taken together can be viewed as forming such an enterprise.

We have argued that this striving for unity is manifest most clearly in the content of embodied, enactive, embedded, extended accounts of cognition, further illustration of which will be given in subsequent sections of this chapter and in chapters 6, 7 and 8. However it is manifest also in the theoretical orientation taken towards this content

which, in various respects and to various degrees, can be regarded as 'dialectical'. We said in chapter 1 that from a Hegelian viewpoint dialectical thought is the style of thinking most suited to the project of unifying subject with object. As the unification of spirit with the world was, on Hegel's account, primarily a rational enterprise anyway – a process whereby the world is finally fully comprehended – dialectical thought could be seen as sufficient in itself to bring about the end of alienation. It is a matter of debate (see chap.2) to what extent Marx's concretizing strategy was in conflict with this i.e. to what extent a Marxian account should be regarded as placing dialectics in the spheres of practical action and/or material phenomena, and how far this deviates from the spirit of Hegel's own account. Nevertheless we might at least say that a dialectical account of the predicament of the human agent was, for Marx, a prerequisite for any satisfactory theoretical understanding of how alienation might be ended – even if that end itself could only be brought about through action. Thus on both accounts the project of the unification of the human agent was inextricably linked to a unifying mode of theorizing i.e. to a 'dialectical' approach.

The dialectical dimension of modern day cognitive science reveals itself in two ways. One is that whereby the attempt to achieve unification on different fronts (i.e. the attempt to unify the human subject with different objects) is accompanied in each case by the contestation of the legitimacy of particular dichotomous categories. Thus, as we have shown above, on one reading at least 'embodied' accounts aim to resolve the dichotomy between mind and body, enactive accounts aim to resolve the dichotomy between cognition and action and embedded/extended accounts aim to resolve the dichotomy between agent and world. We say 'on one reading' here not because there is an alternative reading for which the dichotomy resolving interpretation is less appropriate, but because there is a certain amount of flexibility about how terms like 'embodied', 'enactive' etc. are thought to map onto different theoretical perspectives (see e.g. Wilson, 2002; Anderson, 2003; Ziemke, 2003; Torrance, 2005; Goldman & De Vignemont, 2009).

The second way in which the new cognitive science specifically engages with dialectical themes is that in which some perspectives incorporate a number of elements associated with a dialectical outlook. We will examine this in more detail when we look

at 'Enactivism and the Embodied Mind' (next chapter) and 'Dynamics and Emergence' (see below).

A Unifying Concept

In suggesting that a move to unification (and away from alienation) informs the trajectory of recent cognitive science the intention, in part, is to provide an answer to the question of what it is that the perspectives variously labelled as 'embodied', 'extended', 'embedded' and 'enactive' (as well as 'situated', 'dynamic' etc.) have in common. Without the unifying motif of "unification" itself there is arguably no necessary connection between e.g. an argument for the neurological basis of empathy (Gallese) , the hypothesis that emotions may play a part in rational deliberation (Damasio), and the characterization of a certain type of action as 'epistemic' (Kirsh & Maglio). Clark's attempt to articulate a commonalty in terms of "an escalating retreat from the inner symbol" (Clark, 2001, p.120) whilst useful in many respects, does not account for the breadth of the movement and excludes significant contributions made by those who see no problem with inner symbols (e.g. Damasio, and to some extent Gallese).

Meanwhile analytical accounts such as Wilson's (2002) take us in the opposite direction. Although Wilson declares (perhaps ironically) that "a movement is afoot in cognitive science" (p.623) she alights on just one facet of that movement ("embodied cognition") and then proceeds to dissect it further into six distinct claims, most of which *when analysed in isolation* do not seem defensible. This fragmentary strategy, which has also been adopted by others (e.g. Ziemke, 2003), whilst again undoubtedly useful in some respects, takes us further away from any sense of a movement with a single underlying thread. Our objection here is perhaps similar to Hegel's objection to the modus operandi of the 'Understanding' (Hegel, 1817a, p.113) or Engels' objection to a 'metaphysical' approach (Engels, 1887. p.34). In considering the components of a whole in separation from, and in contradistinction to each other, the cause of analytical precision is advanced, but a price is also paid insofar as an understanding of the relatedness of the component parts is lost.

Currently, then, the theorist who wants to characterize the movement in cognitive science is forced either to do so by listing a long string of adjectives ('embedded', 'embodied' etc…) or must try to subsume as many features as possible under one heading (e.g. 'situated') and jettison those aspects of the movement which can't be made to fit even on the broadest definition (e.g. Robbins and Aydede, 2009, pp.3-10).Whilst it is not our contention that a phrase such as 'unalienated cognition' or 'unified cognition' can act as a replacement for any of these terms, it *is* our contention that an understanding of the movement as one which is towards unification (and away from alienation) can help to highlight features and connections that would otherwise remain obscure.

An Historical Dialectic?

At this point the question might arise of whether this movement towards unification and away from alienation can itself be understood dialectically. Is there - as we suggested was possible in the previous chapter - an 'historical dialectic' at work in cognitive science, a dialectic which expresses itself through the transformations which cognitive science has undergone in its development over time?

Whilst such an idea sounds appealing, elaborating on it in any detail seems problematic. We could perhaps characterize the shift of perspectives from traditional cognitive science to modern cognitive science as a 'crisis' or a 'revolution' (Newell, 1995; Boden, 2006; Di Paolo et al, 2010). We might then see in this fact confirmation of the dialectical view that change over time (in thought or nature) is not a smooth, continuous process but one which involves "qualitative leaps" (Engels, 1887, p.67; Cornforth, 1952, p.61), an idea which, on Engels' analysis at least, seems to be linked with the law of the transformation of quantity into quality. The problem with such an account would seem to be that by itself it is not really telling us very much, or anything very distinct from other, albeit much later, accounts of scientific revolution such as that provided by Kuhn (1962).

Alternatively, again following Engels' lead, we could perhaps see the revolution in cognitive science as indicative of a 'negation' of the previous state of affairs, with

current cognitive science offering an anti-thesis to the thesis of traditional cognitive science. It is not clear from Engels' own account whether 'qualitative leaps' and 'negations' can co refer or whether we have to choose between these two analyses in any particular instance. We might note however that the latter may offer difficulties for our own perspective. If current cognitive science is the 'negation' of previous cognitive science this suggests the immanence of a further 'negation of the negation', which in turn implies the one-sidedness of current cognitive science. Our aim, however, is to show that modern cognitive science is in 'retreat from alienation', a position which is somewhat undermined if the totality of its current outlook is reduced to one side of a dialectical coin.

We could perhaps surmount this problem by changing our time scale slightly. The relevant 'dialectic of mind' we might suggest, did not begin with cognitive science but rather began at some earlier point, for example with behaviourism. Here we might note Newell's own observations on the relation of early cognitive science to behaviourism. He (and Herbert Simon), he suggests, were not motivated "to react against behaviourism" because they were not "deeply involved in psychology" (Newell, 1995, p.150). Nevertheless he allows that "a scientific group has as its principle the reaction against the previous one – in a dialectic way" and that, as an example of this, the revolution in cognitive science was "essentially a revolution against behaviourism" (ibid).

Traditional cognitive science, with its 'cognitivist' outlook, then, could be seen as the negation of behaviourism – insofar as it emphasized the importance of internal symbolic processes to an explanation of human intelligent action. The arc would then be completed by recent cognitive science, playing the role of the 'negation of the negation', reasserting the importance of bodily action, and thus signalling "a return to the beginning but at a higher level of development" (Cornforth, 1952, p.112).

However, whilst much of interest could be derived from such an analysis there is also surely a degree of arbitrariness about it. We have shifted the starting point of the dialectic to suit a particular explanatory purpose, but we cannot justify such a starting point independently. We could, for example, just as easily have started the 'dialectic of mind' with 'introspectionism' or 'phrenology'. Moreover, for the sake of theoretical

elegance we have posited cognitivism as the negation of behaviourism and contemporary embodied cognitive science as the negation of cognitivism, but there are likely to be many ways in which the relationship between these approaches does not conform to, and is not well explained by, such a model. [10] Such problems, we might observe, seem endemic to historical conceptions of a dialectic rooted in the rather vague notion of the 'negation of the negation'.

Nevertheless, we should perhaps not take this vagueness as an indication that there are not patterns present in the historical unfolding of cognitive science which cannot be fruitfully considered in dialectical terms. We might argue, for example, that there is at least a basic dialectic of sorts already implicit in our notion of a 'retreat from alienation'. In chapter 4 we dismissed as fanciful the idea that the trajectory of cognitive science could itself be recast as that of Mind coming to know itself. However, even if we attach little relative significance to the strivings of cognitive science on the stage of world history, we can still characterize it as a subject with a mind (constituted by the minds of various cognitive scientists), which also takes mind as its object. On this basis if traditional cognitive science shows this mind thinking about itself in a narrowly mentalistic way, then contemporary cognitive science shows this mind rediscovering itself in the sensuous physical world. That which had previously seemed other (or at least peripheral) to it – the physical body and its activity, the physical world, other people – has now become reincorporated as part of it.

Here, then, the retreat from alienation *within* cognitive science can also be recast as the retreat from alienation *of* cognitive science. The movement of cognitive science is the movement of its (collective) mind freeing itself from alienation by recognizing itself in the objectivity of the world. Such an account would seem to have some descriptive value provided that one does not take too literally the personification of cognitive science involved. It should be noted however that it is not central to the perspective of this thesis.

---

[10] For example, 'functionalism' – which has many affinities with a classical cognitive scientific outlook – might on one analysis be seen as a synthesis of behaviourism and identity theory.

Damasio & Noë – Anti-Dichotomous Theorists

Returning to the retreat from alienation *within* cognitive science – we have so far said that this can be further characterized as involving a striving for unification, and that the pursuance of such a project is sometimes accompanied by additional dialectical elements. In the remainder of this chapter (and in later ones), we will look in greater detail at particular examples of these unifying approaches within contemporary cognitive science.

In this section we will consider briefly the accounts of two theorists, Antonio Damasio (1994, 2000, 2003) and Alva Noë (O'Regan & Noë, 2001; Noë, 2002, 2004), both of which can be considered as 'anti-dichotomous'. Damasio's account goes some way toward confronting the dichotomy between 'reason' and emotion, whilst Noë's confronts that between perception and action. These case studies are chosen as much for their differences as similarities. We might say that if there were a continuum running from 'alienated perspectives' to 'unified perspectives' then Damasio's account would be significantly further away from the 'unified' end than Noë's. We will begin with Damasio.

Damasio

Damasio's account might be taken as belonging to the 'embodied' camp insofar as the dichotomy it deals with can be regarded as a subset of the more general 'mind/body' dichotomy. Damasio is specifically interested in the traditional separation between 'reason' and 'emotion' where the latter is interpreted in terms of bodily events. However his unificatory strategy is the reverse of that of cognitivist theories which seek to establish that "emotions essentially involve cognitions" (Prinz, 2005, p.10); instead he wishes to demonstrate that cognition is to some degree facilitated by emotion. His argument runs roughly as follows: We often find ourselves, he says, in a situation where we have to make a decision and where there are a range of possible practical outcomes depending on which decision we make. Furthermore such situations are often

time constrained – we need to make a decision *fast*. Damasio gives the following, very un-Marxian, example:

> Imagine yourself as the owner of a large business, faced with the prospect of meeting or not with a possible client who can bring valuable business but also happens to be the arch enemy of your best friend, and proceeding or not with a particular deal. (Damasio, 1994, p.170)

In such situations we may not have the time or the memory capacity to make a decision purely on rational grounds; we can't, for example, do a cost benefit analysis which takes into account all the possible consequences of a decision one way or the other and the probabilities of their occurrence:

> At best, your decision will take an inordinately long time, far more than acceptable if you are to get anything else done that day. At worst, you may not even end up with a decision at all because you will get lost in the byways of calculation. Why? Because it will not be easy to hold in memory the many ledgers of losses and gains that you need to consult for your comparisons   (ibid. p.172)

Yet typically we *are* able to make such decisions, and often with successful outcomes - the question is how? According to Damasio, when considering a range of options certain of these options (or images of the outcomes of these options) are accompanied by negative or positive bodily feelings which help to guide the individual's choice. These bodily feelings are the 'somatic markers'.

In trying to understand exactly what a 'somatic marker' is it is useful to consider this account as in many respects an attempt to make respectable the commonplace observation that we often act 'on gut feeling' (Damasio, 2003, p.147). However it should perhaps be noted that the 'bodily feelings' which constitute somatic markers are a little out of sync with our everyday conceptions of feelings insofar as it is in no way necessary to the existence of the former that they are *felt*. Somatic markers, says Damasio, can "operate covertly" and "do not need to be perceived consciously." (Damasio, 1994, p.180).

We will not here look closely at the question of the origins of these entities. We will merely note that although Damasio seems to suggest some primitive grounding in

primary emotions, he also seems to think that the form they take can largely be explained by "education and socialisation" (ibid. p.177). What we *do* want to draw attention to, however, is (what we might call) the *logical* mechanics of the story. Damasio is very clear that the somatic marker does not do our thinking for us. Rather it intervenes at crucial moments to cut down the number of choices we have to act upon. This 'pruning' mechanism allows us to carry on with our rational deliberation with a greater chance of success:

> The automated signal protects you against future losses, without further ado, and then allows you to choose from among fewer alternatives. There is still room for using cost/benefit analysis and proper deductive competence, but only after the automated step drastically reduces the number of options…Somatic markers do not deliberate for us. They assist the deliberation by highlighting some options (either dangerous or favourable) and eliminating them rapidly from subsequent consideration. (Damasio, 1994, pp.173-174)

Postponing, momentarily, criticism of the details of this story we might say that Damasio's undoubted achievement is to make credible the idea that a neurological/biological story *can* be told which might undermine the conception of reason as essentially separate from embodied emotion. From the point of view of the 'retreat from alienation' his account is both in keeping with the spirit of the project of human unification and adopts an anti-dichotomous strategy to achieve this end – although in both cases to a very limited extent. Insofar as this is the case it also appears to be to some degree consistent with one of Hegel's own observations on the subject of 'feeling' - that:

> The difficulty for the logical intellect consists in throwing off the separation it has arbitrarily imposed between the several faculties of feeling and thinking mind… (Hegel, 1817c, p.93).

Nevertheless there are problems with Damasio's account. Perhaps surprisingly, given the subject matter of this thesis, these problems are not primarily connected with his focus on business scenarios or his reduction of rational thought to cost-benefit analysis. Although a case can be made that his market-oriented outlook does have some strong relation to the more abstract features of his argument, we will not pursue this line of reasoning here. Instead we will concentrate on those abstract features themselves. Below I will outline two of the problems that Damasio's account seems to present. This

will, on the face of it, involve a substantial departure from any Marxian related considerations. However I hope to then show that these problems are ultimately amenable to a Marxian analysis (even if a rather sketchy one.)

Problem 1

One initial response to Damasio's account might itself be cast in terms of a 'gut feeling'. Anyone with a modicum of sympathy for the approach of the later Wittgenstein might well feel some unease at Damasio's introduction of a hypothetical inner entity/process, not necessarily accessible to consciousness, whose main role is to help in the completion of an otherwise seemingly intractable cognitive task. There is a characteristic argumentative structure used by Wittgenstein (1953) which might be labelled 'the transitive argument'. This argument is of the form: 'You say that a system's ability to get from A to C can be explained in terms of an intermediary stage where it goes from A to B, but in your example you are presupposing in this ability to get from A to B, the self same skills/knowledge that were required to get from A to C in the first place."

An example of this transitive argument is found in his treatment of the question "How does an individual know what colour to pick when he hears the word 'red'?" Wittgenstein imagines someone replying that an individual will know what colour to pick because he can pick the colour "whose image occurs to him when he hears the word". Wittgenstein's response to this is:

> But how is he to know which colour it is whose image occurs to him'? Is a further
> criterion needed for that?   (Wittgenstein, 1953, p.88)

In other words the 'inner image' has no role to play. That which would allow me to unfailingly identify an inner image as red ought also to allow me to unfailingly identify a red object as red without the help of the inner image. Therefore it makes no sense to say that I typically identify red objects by means of a red inner image. (Here we are reminded of Brooks' remark that "it turns out to be better to let the world serve as its own model" – 1991b, p.396). This argumentative structure, we might note, has also

been used more recently by Dennett in relation to a debate about the ability of the brain to 'fill in' detail of the world which is not provided by our perceptual apparatus. Noë summarises Dennett's argument as "If the brain knows what it needs to fill in, then for whose benefit is the operation of filling in performed?" (Noë, 2002, p.3)

A question that might concern us, then, is whether Damasio's somatic marker argument might also be susceptible to a similar 'transitive argument'. Is the somatic marker "a wheel that can be turned though nothing else moves with it?'(Wittgenstein, 1953, p.95) i.e. an explanatorily useless part of any hypothesis?

In fact it is pursuit of an adequate formulation of a transitive argument in this context that exposes one of the weaknesses of Damasio's account. If we take the 'colour' argument (above) as saying : "If I know how to identify an 'inner' red then I should know how to identify an 'outer' red directly" then an equivalent for the somatic marker hypothesis would seem to be: " If I know how to use somatic markers to prune my choices then I should be able to circumvent that process by not considering certain choices in the first place". Here however we immediately notice a problem in the use of the phrase "If I know…" A supporter of somatic markers might quite rightly point out that the use of this phrase exposes a flaw in the reasoning. Somatic markers are not the result of my 'knowing' anything; they do not result from any kind of strategy employed at the personal level. Rather the somatic marker just makes its presence felt (or unfelt) and I am the beneficiary. The somatic marker precisely *is* able to 'circumvent the process' but this does not mean that 'I' am able to. In response to this we must therefore amend our original formulation to eliminate mention of knowing selves. Suppose we alter it to "If *the system* knows where to use somatic markers *it* should know how to restrict choices without using them"? Does this argument work?

Again, a defender of Damasio's account might here claim that somatic markers are not merely distinct from the strategies of selves but distinct from 'the rational strategy of the system'. In a way this seems like it must be trivially true. The whole point of the hypothesis is to show that merely rational means cannot achieve the desired ends so some other element must be involved in our decision making processes. This other

element is by definition 'not rational', so if by 'distinct from the rational strategy of the system' we just mean 'not rational' there cannot be any argument here. However it seems on this limited interpretation we are still free to interpret the somatic marker as being part of the 'deliberative strategy' of the system – just not the *rational* strategy. What this might mean is not entirely clear but we might assume for instance it means that we could still model the system with somatic marker included. In this case we can still ask "If the system has the wherewithal (however characterised) to apply somatic markers to particular choices –why need it throw up those choices in the first place?"

If a defender of the somatic marker hypothesis is to attempt any kind of counter-argument to this it seems that they could only do so by excluding the somatic marker from 'the deliberative system' altogether – and indeed this seems to be implied by Damasio's comment that "somatic markers do not deliberate for us" (Damasio, 1994, p.174) On this view the somatic marker would be more like an environmental variable, acting blindly and external to the system. Choices are pruned 'out there' (just as the wind might prune the branches of a tree) and the system then gets to work on the result. This version of things would seem to defeat the transitive argument, but at what price? We now have a complete separation of emotional and rational processes – with 'reason' clearly at the centre of things once more. Any attempt at unification has gone out of the window.

Problem 2

Moreover this division of labour between the work of the somatic marker and that of the deliberative system makes problematic the account of how the rational system gets its job done. Damasio suggests that the somatic marker had to be called in because of our limited time resources and memory capacity, the suggestion being that the somatic marker does its work by reducing a large amount of choices (and hence decision paths to consider) to a smaller more manageable amount. Thus the problem is presented as a quantitative one, but once the problem is presented in this way there seems to be

something arbitrary and unconvincing in the idea that a non-rational mechanism for reducing the amount of choices will be able to provide a solution. Is there a third mechanism, for example, which guarantees that the somatic marker will prune away just enough choices every time to make the operations of the rational agent feasible?

One suspects, in fact, that the problem Damasio's account is circling around is not a quantitative one at all but the *in principle* one of how a purely rational agent is able to decide on a course of action in a world of unending possibilities. The difficulty we are alluding to here is similar to that which has surfaced in AI as 'the frame problem' (McCarthy & Hayes, 1969; Dennett, 1984) and in Philosophy as 'the rule following paradox' (Wittgenstein, 1953; Kripke, 1982). Like the agent in those two scenarios Damasio's decision-maker seems to be faced with great difficulties in answering the question 'How do I know what to do next?' The rational resources that she has at her disposal (facts, rules, inferential mechanisms etc.) are not sufficient to determine an outcome in any particular situation. To his credit Damasio therefore looks elsewhere – to non-rational resources – for a solution. However, as these non-rational resources are only given a quantitative role to play – the reduction of the number of choices – it is hard to see how they can impact significantly on the qualitative problem of 'choosing' itself.

This difficulty is further highlighted by Damasio's characterization of the decision making process as one which involves selection from a range of 'images'. In deciding whether or not to do business with the arch enemy of our best friend, for example, we are faced with a number of representations of future outcomes:

> Examples of what the images would depict include meeting the prospective client, being seen in the client's company by your best friend and placing the friendship in jeopardy; not meeting the client; losing good business but safeguarding the valuable friendship, and so forth. (Damasio, 1994, p.170)

Let us allow that this description of things makes sense, that there are a finite amount of such outcome images and that the somatic marker can effectively prune away the majority of them (which may be allowing too much). Suppose then we imagine an

ensemble of negative somatic markers attaching to all future outcome images except one – perhaps we only feel good (or don't feel bad) about an image which somehow illustrates continued friendship (an image of my friend and I having a drink together?). Where does that leave the rational system? Is what it is to do next determined by the selection of this image? It is difficult to see how this could be the case, for, again in Wittgensteinian fashion we might point out that the image does not bring with it a set of rules determining how it is to be applied in all future circumstances. Moreover, even if it did, we might need rules to interpret these rules and so on (Wittgenstein, 1953, pp.75-88). In short no amount of image pruning by the somatic marker need do anything to lessen the interminable amount of decisions faced by a purely rational system.

A dialectical solution?

We have highlighted, then two possible problems with the somatic marker hypothesis. Firstly, if it manages to evade the 'transitive' argument it only does so by isolating in a problematic fashion the 'somatic' element of decision making. Secondly, insofar as there is a problem with rational decision making it seems unlikely that such a problem could be solved by a quantitative 'pruning' of choices.

Both of these problems, we might argue, stem (in part) from the separation that still persists between the rational and the somatic in Damasio's account. Whilst the incorporation of a somatic element with a role to play in deliberation is a positive move, the restriction of the somatic and rational to their own isolated arenas of operation is less so. In dialectical terms there is little exploration of ways that these two components might "interpenetrate" or form a reciprocal unity. The attempt to "break up the rigidity" (Hegel, 1817a, p.53) between feeling and thinking is not thoroughgoing enough.

Of course, Damasio might well suggest that the examples he gave were deliberately simplified and that in any real life situation there is always a to–ing and fro–ing back and forth between rational and somatic components – a little theorizing, followed by a bit of pruning, then back to theorizing again and so on. In this way he might argue that

he too believes that the somatic and the rational are inextricably 'mixed up' together. However from our point of view this would still be a relatively un-dialectical kind of mixing up. Like Merleau-Ponty's conception of 'mechanical action' it reduces the interaction between emotion and reason to "a series of uni-directional determinations" (Merleau-Ponty, 1942, p.161).

A more unified account, we might speculate, would not be one where the agent, as identified with a rational/deliberative system, is influenced from without by a somatic mechanism. Instead the deliberative system itself would be seen as incorporating the 'rational' and 'emotional' in a more integral way, with the notion of a 'purely rational' choice being revealed as something of an abstraction.[11]

Such an account might go some way towards offering a solution to the problems outlined above. If the somatic marker – or however it might need to be re-characterized in a unified system – were fully integrated with the process of rational decision making i.e. if 'choosing' itself were viewed as a partially somatic/affective/emotional process, then the 'transitive argument' could no longer apply. There would no longer be a confusing dichotomy between the 'deliberative system' and some other part of the mechanism which appears to be deliberating but isn't, and so we would no longer feel compelled to question the role of the latter. Likewise, the rejection of the idea that a "purely rational agent" has to confront the world alone at any point in the proceedings helps us bypass the difficulties which externalized somatic pruning was designed to solve but would be unable to solve i.e. frame problem/rule following type problems which follow on from a purely rational conception of decision making.

---

[11] This is hardly a new idea of course. Ron Chrisley, for example, argues that "The notion of a 'purely' rational AI system is a distraction. It's questionable whether any implemented AI system has ever been purely rational, or whether such would even be possible."(2003, p.138) He seems however to be making a slightly different point to me. His suggestion, as far as I understand, is that it is hard to imagine an embodied system where every significant physical difference gives rise to a rational difference, whereas the point here is that outside the realm of artificial agents it may not always be useful to consider rational and non-rational processes in separation.

These remarks are, of course, necessarily vague and impressionistic. It might be possible to put more meat on the bones of our analysis both by more detailed speculation and by reference to existing accounts which *have* attempted to unify emotion and reason in a more fundamental way, either in the context of a general philosophical holism (e.g. Dreyfus, 1972) or in relation to neurological research (e.g. Hardcastle, 1999). Our aim here, however, is not to put forward or review theories of the relationship between emotion and cognition in any depth but rather to present Damasio's account as an example of a weak attempt at unification – a case wherein an author has successfully identified a tension within a particular dichotomy but has made no sustained attempt to resolve that dichotomy.

Noë

In this respect Damasio's account is to be contrasted with that of Noë (O'Regan & Noë, 2001; Noë, 2002; Noë, 2004) – for Noë does try to confront a dichotomy head on and the results are more far-reaching as a consequence.

Although Noë's work covers a lot of ground one of his main targets is the "classical sandwich" (Hurley, 2001) model of cognition described briefly in our previous chapter. According to this model perception and action stood at opposite ends of a linear sequence separated by cognition. Noë's "enactive approach" (Noë, 2004, p.2) challenges this serial conception of things by questioning the separateness of action and perception. Influenced greatly by Merleau-Ponty's (1945) own active account of perception, he argues that perception is not separate from action but rather is a type of action:

> Perceiving is a way of acting. Perception is not something that happens to us, or in us. It is something we do. (Noë, 2004, p.1)

Central to this account of perception are the notions of 'sensorimotor contingencies' and 'sensorimotor knowledge'. What enables coherent perceptual experience in relation to a particular object is the distinct pattern of sensory changes that arise from the perceptual probing of that object with a particular sensory modality. Such patterns

constitute the 'sensorimotor contingencies' and (practical) knowledge of them is prerequisite for meaningful perceptual content (O'Regan & Noë, 2001; Noë, 2004).

We can already see, then, that Noë goes further than Damasio in confronting the dichotomy at the centre of his investigation. Whereas the latter had posited a mere external relatedness between emotion and reason Noë, in a more Hegelian spirit, wants to unify perception and action in a principled way. His 'constitutive claim' is not just that perception is bound up with action but that perception is a kind of action. We should note here how he explicitly wants to put some ground between his own account and other accounts in this respect:

> Most recent work on the relation of perception and action stop short of making the constitutive claim that defines the enactive standpoint. It does not treat perception as a kind of action or skilful activity (or as drawing on a kind of sensorimotor knowledge), rather it treats (a good deal of) perception as for the guidance of action. (Noë, 2004, p.18)

In fact this distinction between teleology and constitutiveness is perhaps both a weakness and a strength in Noë's account. That he feels the need to exclude 'perception for action' from his own version of enactivism suggests a certain level of fragmentedness in his approach – for it is surely the case that 'perception for action' and 'perception as action' are related to some degree. (Similar considerations seem to motivate Torrance's own observations concerning the lack of teleology in O'Regan & Noë's brand of enactivism – Torrance, 2005, p.7). Nevertheless that Noë wishes to posit an essential unity *of some kind* between perception and action counts as a tick in the dialectical box.

Moreover, this unity between perception and action also recalls some of the features we grouped under the heading of 'praxis' in chapter 2. This is perhaps no accident since, as we have said, Noë's account is to some degree influenced by Merleau-Ponty's work. It will be remembered that Merleau-Ponty's own account of 'active perception' could in turn be convincingly related to Marx's conception of the active relationship of the agent to the environment, and in particular to his critique of Feuerbach's materialism according to which:

> …the thing, reality, sensuousness, is conceived of only in the form of the object, or of contemplation, but not as sensuous human activity, practice (Marx, 1970a, p.121)

Like Marx and Merleau-Ponty, then, Noë's account of perception as skilful activity also emphasizes an active relationship to the world and so is at odds with classical materialism which "presents the world as if it were a merely passive object of perception" (Avineri, 1970, p.75).

There is more of interest to us in Noë's enactive account, however, than the simple fact that he characterizes perception as a type of action. There is the additional fact that he uses this position to tackle a number of related puzzles of perception. For example the 'grand illusion' argument (which Noë credits to Dennett, 1991, and Blackmore et al. 1995) – that phenomena such as change blindness and blindspots indicate that visual perception does not give us instantaneous access to a detailed and complete picture of the world, and that therefore that "we are mistaken in our assessment of how things seem to us to be" (Noë, 2004, p.54) – is shown to arise from a mistaken 'snapshot' conception of visual experience. It is not that we are deluded into thinking that we have better visual access to the world than we really have. In fact we do have full access to all the detail of the world but such access is not to be understood on a model according to which the agent is a passive spectator to a static given (nor, says Noë, is this the layman's conception of things anyway). Rather we have such access to the world over time as active explorers of that world:

> My visual world is not the field available to the fixed gaze. The visual field rather is made available by looking around. (ibid. p.57)

Likewise Noë uses his sensorimotor approach against more traditional philosophical problems of perception. One such is the 'argument from illusion' (see e.g. Ayer, 1956). According to this argument cases of non-veridical perception are taken as evidence that we never have direct perceptual access to the world. A plate, for example, may look small and elliptical from a certain distance and angle, when it is actually large and circular. The 'look' of the plate must therefore be something separate from the plate itself for it presents the plate as possessing properties that it does not in fact have. Such 'looks', it is argued, are instances of 'sense data' and it is with them – rather than the world itself - that we have direct perceptual acquaintance.

From Noë's point of view this line of reasoning is misguided. The fact that objects change in appearance as we explore them perceptually does not warrant the conclusion that we do not have access to objects as they actually are. Rather it is through this pattern of sensory changes that we ascertain the actual size and shape of the object:

> It is in the possible changes in P shape that the real shape is encountered.   (Noë, 2004, p.86)

We expect (or discover) a circular object to look elliptical to a greater or lesser degree from different angles for these are the sensory patterns typical of a circular object: "We see its circularity *in* the fact that it looks elliptical from here" (ibid p.84). There is therefore no distinction to be made between the object as it appears to us and the object as it actually is, and in this sense perspectival properties "are perfectly 'real' or 'objective'" (ibid. p.83).

We have said already that Noë's account of perception is praxis-like in its emphasis on an active orientation to the object. Here his use of an enactivist position to provide solutions to philosophical puzzles recalls further elements of Marx's philosophy of practice. It will be remembered that Marx's account included not only ideas about the active practical nature of man, but also theses concerning the relationship between practical activity and philosophy. This is one respect, we might say, in which a Marxian approach differs from a merely phenomenological one (such as a 'skills' oriented Heideggerian account e.g. Dreyfus, 1972; Wheeler, 2005). Practical activity is seen not just as a crucial category for understanding human 'being-in-the-world' but as a source of solutions to philosophical problems (and ultimately as a means of dispensing with philosophy altogether). This was one of the implications of Marx's assertion that:

> All mysteries which lead theory to mystification find their rational solution in human practice and in the comprehension of this practice. (Marx, 1970a, p.122)

Taken in conjunction with Marx's emphasis on our active relationship to the world and its contents, this position has the following consequence: philosophical problems which arise from a passive conception of our relationship to the world dissolve once one has a fuller comprehension of our active relationship to the world – and in some instance solutions to these problems are brought about by action in the world.

We can see how this applies to Noë's account. Noë does not advocate action itself as a means of superceding the fragmented outlook of detached theory – for this we have to turn to Varela et al. (1991 - see next chapter). However he does use an active conception of our relationship to the world to dissolve puzzles that had previously arisen as a result of a passive, static conception of our relationship to the world. The 'mystifications' implicit in the 'grand illusion' argument and the 'argument from illusion' are thereby revealed for what they are – the by-products of an outlook that failed to view things "in their own being and movement" (Hegel, 1817a, p.117) and which failed to characterize human beings in accordance with their "real, active" nature (Marx & Engels, 1970b, p.47). The scenarios they present are precisely of the sort that Marx would have labeled 'abstract' insofar as they leave out all that is essential about our ongoing interaction with the world and posit instead a picture of the agent as static receptacle of sense impressions. Noë overturns this picture – replacing man as a passive spectator with man as active producer of his own perceptions.

Intersubjectivity & Mirror Neurons

Damasio's and Noë's accounts, then, are both attempts to tackle dichotomies which militate against the unification of the agent. In Damasio's case the dichotomy was that between emotion and reason and in Noë's case it was that between action and perception. Noë's account was both more successful in this respect and incorporated additional Marxian elements. We might now turn to another unifying thread in current cognitive science – its concern with intersubjectivity – which can also be expressed in terms of the resolution of a dichotomy, the dichotomy between self and other.

We saw in Chapter 4 how traditional cognitive science had an individualistic outlook on mind and cognition, and that this had consequences particularly for its approach to the question of 'other minds'. In an epistemological equivalent to Marx's 'alienation of man from man' the relationship of individuals to each other was viewed as problematic and the gulf that separated individual minds was seen as bridgeable only indirectly through inference or some similar theoretical procedure.

More recently many researchers and theorists within cognitive science (e.g. De Jaegher & Di Paolo, 2007; Froese & Di Paolo, 2009; Gallagher, 2005; Gallese 2001, 2003; Thompson, 2001, 2007; Torrance & Froese, 2011; Zahavi, 2001) have focussed on intersubjective or interpersonal dimensions of cognition. In various ways the idea has been expressed that it is unviable to take individual self-contained minds as a starting point or, in some cases, as a given. As Thompson, in a wide ranging survey of the field as it was ten years ago, put it:

> What recent cognitive science has begun to drive home is that the embodied mind is intersubjectively constituted at the most fundamental level. (Thompson, 2001, p.4)

Such approaches bypass the need to explain how the gulf between self and other is bridged for they either deny the existence of such a gap, or at least that it is of the sort imagined by traditional philosophy of mind and theory theorists. As Thompson points out, their grounds for doing so range from purely phenomenological considerations - for example that any kind of sense experience presupposes an 'open intersubjectivity' (ibid. p.15) – to speculative interpretations of research in developmental psychology - e.g. "that human agents possess, at birth, interpersonal body schemas for emotional contagion and facial imitation" (ibid. p.7 - in reference to Meltzoff and Moore, 1999.)

Some later work (De Jaegher & Di Paolo, 2007; Froese & Di Paolo, 2009; Torrance & Froese, 2011) has focussed not just on the primacy of intersubjectivity but on the dynamic characteristics of interactions between individual agents. Such an interaction can have "an autonomy of its own which is separate from the autonomy of the individual participants" (Torrance & Froese, 2011, p.25) and so can facilitate or hinder the actions of individuals irrespective of their intentions (ibid. pp.25-27). This description of 'participatory sense making' (De Jaegher & Di Paolo, 2007) perhaps brings to mind Sohn-Rethel's account of commodity exchange wherein "the action is social, the minds are private" (Sohn-Rethel, 1978, p.29).

However whilst the fact of this intersubjective turn in cognitive science is of interest to us in principle and constitutes one dimension of a 'retreat from alienation', there is perhaps not much to be gained in detailing the myriad ways it has been developed by various researchers and theorists, particularly when such developments may in some

cases have been effectively theorized in relation to non-Marxian forebears such as Husserl and Heidegger (see e.g. Zahavi, 1999; Thompson, 2007; Froese & Di Paolo, 2009). What is of particular relevance to our case is that the turn to intersubjectivity is often accompanied by an awareness of its internal relation to other unifying conceptions of the agent as embodied, enactive etc. Thus Thompson posits a single "enactive" perspective whose elements include 'self-other co-determination', embodiment and emergence (Thompson, 2001, 2007). We will look at other aspects of this version of enactivism in more detail in the next chapter.

Gallese

Also of interest to us are certain accounts which contain additional features particularly relevant to a Marxian perspective. One such is that of Gallese (Gallese 2001, 2003; Gallese & Lakoff, 2005), whose work we shall now briefly examine. Gallese's main interest is in the neurological mirroring systems found in humans and other primates. In the case of macaque monkeys he describes how:

> A particular set of neurons activated during the execution of purposeful, goal related hand actions, such as grasping, holding or manipulating objects, discharge also when the monkey observes similar hand actions performed by another individual. We designated theses neurons as 'mirror neurons'. (Gallese, 2001, p.35)

Roughly equivalent (at least for our purposes) results for human beings are also discussed. These show that humans too have a 'mirror matching system' – motor circuits in the human brain are activated both when an action is observed and when it is performed:

> When we observe goal related behaviours executed with different effectors, different specific sectors of our pre-motor cortex become active. These cortical sectors are those same sectors that are active when we actually perform the same actions. (ibid., p.37)

Gallese sees in this, and other, data evidence of a neurological basis for a "shared meaningful intersubjective space" (Gallese, 2003, p.517) that we all inhabit. There are (we might argue) two components to his argument.

One is the general idea that neurological mirroring of any sort supplies us with an answer to the question of how, on a subpersonal level, it is possible to bridge the gap between 'observer and observed'. If we redescribe the traditional 'problem of other minds' argument in Wittgensteinian terms we might say that it is an instance of the problem of the lack of a 'criterion for the same'. The adherent of analogical reasoning, as exemplified by Theory Theory (Baron Cohen et al, 2000) or the traditional 'argument from analogy' (Ayer, 1954), likes to suggest that the gulf between self and other and can be crossed through the imaginative transposition of first person states to others, but this strategy presents us with a problem: On what basis are we to assume that third person states are 'the same' as first person states? More radically, what can it even mean to say that another person is in 'the same' state as me, when such states are for me, by definition first person states? Wittgenstein expresses this problem as follows:

> If one has to imagine someone else's pain on the model of one's own, this is none too easy a thing to do: for I have to imagine pain which I *do not feel* on the model of the pain which I *do feel*.   (Wittgenstein, 1953, p.101, author's emphasis)

Wittgenstein's solution to this is to deny that, in the typical case, 'mind-reading' (if we are to use this term) is based on inference or the imaginative transference of first person states, and thus that there is a problem of this sort to deal with. Nevertheless, given that we accept that a non-inferential account of intersubjective relations is correct, this still leaves space for a story to be told about the subpersonal processes that enable non-inferential interaction – we still need a subpersonal criterion for 'the same'. It is this, we might say, which neurological mirroring systems provide. They imply a hard-wired means of establishing identity between that which is experienced in the first person and that which is observed in others.

At this general level of abstraction, then, the idea of neurological mirroring might be thought applicable to a broad range of intersubjective encounters including those which are not specifically concerned with the practical engagement of the agent. In this context we note, for example, research into pain-related areas of the brain (specifically the right dorsal anterior cingulated cortex) which indicate common neuronal response to the pricking of a subject's hand and the observation by that subject of another's hand being pricked (Hutchinson et al 1999; Morrison et al, 2004) thereby implying (at least

for Morrison) "a common neural substrate for felt and seen pain" (ibid. p.273). Here it is possible to talk (with or without justification) of a neurological basis for 'empathy' without making any essential reference to the agent's practical involvement in the scenario.

The second component of Gallese's argument, however, goes beyond this, and it is this component which seems particularly amenable to a Marxian interpretation. Although Gallese gives some mention to pain related mirroring (in 2001, 2003) it is not the existence of mirroring systems per se which interests him but specifically mirroring systems located in motor areas, for this implies (for him) an action related mirroring which serves as the foundation for intersubjective relations: "Through a process of 'motor equivalence' a meaningful link between agent and observer is established" (Gallese, 2001, p.42). The suggestion would seem to be that neither inference, nor even a statically conceived 'empathy' – which still reduces agents to the passive roles of 'observer' and 'observed' – is sufficient as a connector. Rather it is through action, and our readiness for it, that we are linked with our fellow human beings at the most fundamental level:

> Action is the 'a priori' principle enabling social bonds to be initially established."
> (Gallese, 2001, p.41)

We have of course to bear in mind here that Gallese's original subjects were not human beings but macaque monkeys, but his point is surely that insofar as an organism is capable of forging "social bonds", such bonds are to be understood as grounded in action rather than in passive cognition.

Such a perspective fits well with Marx's account of the human agent. As we saw in chapter 2, for Marx man was a quintessentially social being, and his social essence was not realized passively but rather through productive activity. The term 'species being' (Marx, 1844a), although used in a variety of ways, was to be partly understood in this connection as denoting man's essence as a socially productive animal. It is in this sense too that the different facets of alienated labour were only separable in the abstract, for alienation from activity and its product goes hand in hand with alienation of man from man. Conversely realization of man's social essence is simultaneously his realization

through "free conscious activity" (Marx, 1844a, p.82). It is through such activity, undertaken in the presence of others, that man confirms his "communal essence". (Marx, 1844b, p.122)

We might say then that Gallese's interpretation of the significance of motor based mirror neurons, whether of not it is warranted by the evidence, is one that is congenial to a Marxian outlook. The suggestion that purposeful action and intersubjectivity form a unity at the most basic (neurological) level is consistent with a theory which asserts that human agents realize themselves as social beings through practical activity. For both Marx and Gallese social bonds are fundamentally practical in nature.

In arguing for a Marx-friendly interpretation of Gallese, however, we do not mean to imply agreement with all aspects of his account. We have, for example, omitted reference to certain details concerning the mechanism by which mirror matching systems are thought to facilitate access to a shared intersubjective space – namely that of 'embodied simulation', which in later work (Gallese & Lakoff, 2005) is held to play a role in understanding and concept formation. This portion of his argument seems problematically representation-oriented in its outlook. Likewise we have also omitted reference to one of Gallese's explanations of the evolutionary function of mirror matching systems – that "social cognition has action control as one of its main purposes, namely controlling the action of others" (Gallese, 2001, p.39). This interpretation of the data seems to incorporate unwarranted assumptions about the inevitability of human competition, somewhat reminiscent of those made in the early stages of Hegel's parable of the 'Lord and Bondsman' (Hegel, 1807, pp.111-119). We will not develop these points further here however.

Objective Dialectics – Dynamics and Emergence

We have so far looked at examples of dichotomy resolution in the accounts of Damasio and Noë, with the latter being characterized as both more thoroughgoing and more praxis-oriented. We have also looked at examples of intersubjective accounts of the

agent, concentrating specifically on Gallese, whose emphasis on a pre-theoretical, action based link between agents mirrors Marx's conception of man as an active creature who realizes his social being through practical activity. We turn now to another significant thread in modern cognitive science, dynamic systems' perspectives. Dynamical accounts (henceforth 'dynamics') are of interest to us because they incorporate some of the dialectical elements we outlined in the work of Hegel and Engels (and later Marxists). In doing so they also present a picture of cognition at odds with the mechanistic 'production line' conception we outlined in chapter 4. There we said that whereas GOFAI emphasized linearity, discreteness and fixity, dialectics emphasizes reciprocity, continuity and holism. These latter elements are also central to dynamics.

The following account is not comprehensive and will not include analysis of many of the concepts one might associate with dynamics (e.g. there is no mention of 'state spaces' or 'attractors'). Instead we will look briefly at a few features of particular relevance to this thesis – 'change, motion and continuity', 'emergence', and 'reciprocal causation' – before moving on to consider how some of these features, and others, are presented in one particular dynamical account (van Gelder & Port, 1995; van Gelder, 1995, 1996).

Change, Motion and Continuity

In order to emphasize the similarities between dynamics and Hegelian-Marxian dialectics it is perhaps useful to begin with the general observation that both dialectics and dynamics are centrally concerned with change and motion. It will be remembered (chapter 3) how, on Engels' account, a dialectical view of the world was one which represented it as in "constant motion, change, transformation, development" (Engels, 1867, p.37). This emphasis on process was meant as an antidote to the 'metaphysical' perspective which viewed the objects of the world "in repose, not in motion, as constants not as essentially variables" (Engels, 1867, p.34). Whilst a metaphysical perspective did have its limited uses, it was the dialectical perspective which grasped the world as it actually was because, amongst other things, it understood that "motion is the mode of existence of matter" (Engels, 1867, p.86). This conception of things was

indebted to Hegel's own account in the *Logic* where there was a similar emphasis on processes in time, as expressed in the idea that 'Becoming' is the "fundamental feature of all existence" (Hegel, 1817a, p.132).

Dynamics is likewise concerned centrally with processes and change, and as with Engels' dialectical outlook, its aim (in the arena of cognitive science) is not just to offer additional insights otherwise neglected by the classical view, but to show that cognition is only fully comprehensible in these terms. As Thelen and Smith put it "there is only process" (Thelen & Smith, 1994, p.39) and later – in a remark reminiscent of Marx's eleventh thesis on Feuerbach – "the central theoretical problem is not stability but change" (ibid, p.42). Such a perspective is also sometimes accompanied by Hegelian terms and references. Thus, for example, Kelso notes that "becoming is a process of change" (Kelso, 1995, p.5) and opens his book with a passage which recalls Heraclitus' maxim that 'it is not possible to step twice into the same river':

> I envisage it [the brain] as a constantly shifting dynamic system, more like the flow of a river in which patterns emerge and disappear, than a static landscape. (ibid, p.1)

One component of this general orientation towards change and process is the idea that 'continuity' is significant to cognition. Dynamicists have a "strong belief in the relevance of continuity to providing accurate descriptions of cognitive systems" (Eliasmith, 2001, p.422). Thus, as with the dialectical critique presented in the previous chapter, dynamical accounts oppose themselves to the discrete mechanistic models of mind as represented by the "sudden jumps or clicks" of discrete state machines (Turing, 1950, p.35; van Gelder & Port, 1995, p.2).

However, in identifying both dialectics and dynamics with a preference for the continuous over the discrete, we are to some degree talking 'abstractly' ourselves. *In the real world* the dichotomy between continuity and discreteness is not fixed. Discrete systems can be understood dynamically – thus for example, "the symbolicist Turing Machine *is* a dynamical system" (Eliasmith, 1996, p.456). Likewise, as we saw in chapter 1, Hegel's conception of a dialectical relation between quantity and quality is one which envisages relatively discrete states "under the figure of a nodal (knotted) line" (Hegel, 1817a, p.160) emerging out of a process of continuous change. The point

however is that the discreteness posited by traditional GOFAI mechanisms is *not* of the real world but is such as can only exist in abstraction, or in treating concrete devices abstractly. It has been part of the role of dynamics to point this out.

Internally related to this hostility to modelling mind as a series of discrete state transitions is an emphasis on the importance of time. There are various ways in which one might understand the significance of temporality to dynamics (Clark, 1998b), but on one reading the emphasis on time is just the flipside of the emphasis on continuity. Understanding a process in continuous real time, rather than the "arbitrary 'step' time" (Port & van Gelder, p.2) of discrete machines, allows one to consider it in the entirety of its movement, rather than as a series of fragmented moments presented in isolation from each other. In this way, for dynamics as for Hegel, it is the transition "which is the essential point" (Hanna, 1996, p.228).

Emergence

A second concept linked with dynamical accounts is that of 'emergence'. The term 'emergence' – used to characterize phenomena not too dissimilar to those which interest modern cognitive scientists - has an independent history which can be traced back at least to the eighteenth century (Corning, 2002). It is also perhaps as central to enactivism - which we will look at in the next chapter - as it is to dynamics. Nevertheless we are justified to some degree in discussing it here as part of a dynamics perspective, for the two are held to be intimately related on many accounts (e.g. Varela et al, 1991; Thelen & Smith 1994; Kelso, 1995; Clark, 1997; Walmsley, 2003).

Varela et al. note that "there is no unified formal theory of emergent properties" (1991, p.88) and to this we might add that even if we jettison requirements for formalism the term 'emergence' seems to be deployed in multifarious ways. Of course we expect there to be diversity both in the types of properties/behaviours susceptible to explanation in terms of the concept of emergence – e.g. 'mind' (Minsky, 1986), behaviour warranting the application of intentional terms (Dennett, 2003), symbolic computation (Varela et al., 1991; van Gelder 1996) – and in the types of dynamic substrate thought capable of supporting emergence – e.g. finger movements (Kelso,

1995), limb movements, (Thelen and Smith, 1996), neurons (Varela et al., 1991), people (Minsky, 1986) and cellular automata (Dennett, 2003). It is in its breadth of application that the usefulness of the concept resides. However some differences of application also imply a difference in concept.

In this connection we might note that Brooks is prepared to use the term 'emergence' to explain the apparent intelligence of both traditional and 'behaviour based' AI systems. In each case "intelligence emerges from the interaction of the components of the system" (Brooks, 1991b, p.418). However Brooks goes on to make the point that the way it emerges in both cases is "quite different" (ibid. p.419). The nature of this difference is examined by Clark (1997, 2001) who argues that a crucial distinction must be made between 'componential explanation' and 'emergent explanation'. In the former "we explain the capacities of the overall system by averting to the capacities and roles of its components, and the way they interrelate" (Clark, 1997, p.104). This mode of explanation is appropriate for explaining "the workings of a car, a television set, or a washing machine" (ibid) as well as the workings of a classical modular AI program. Truly emergent explanations, however, seem to involve more than this on most accounts.

In pursuance of a correct account of genuine emergence Clark notes that definitions of emergent phenomena in terms of processes which are "unexpected' or "unpredictable", are unsatisfactory, for such concepts, are subjective – "what is unexpected to one person may be just what someone else predicts" (Clark, 1997, p.109). Likewise, says Clark, definitions in terms of "controlled" and "uncontrolled" variables (Steels, 1994) are also problematic. On such an analysis emergent properties are those which are not controlled by any simple input parameters to the system ("controlled variables") and which can thus be viewed as "uncontrolled variables". The problem here, according to Clark, is that many classic examples of emergent phenomena (such as Kelso's convection rolls in boiling oil, see below) *could* be tracked to changes in simple, isolatable input parameters (in this case, variations in temperature.) Clark thus plumps for an alternative definition – "a phenomenon is emergent if it is best understood by attention to the changing values of a collective variable" (Clark, 1997, p.112). However, he admits that even this characterization is "weakly observer dependent" (ibid. p.113).

Contra Clark, Walsmsley (2003) has suggested that reference to unpredictability needs to be kept, but must be squared with the fact that the covering laws of dynamic systems theory imply the predictability of the phenomena they describe. He thus makes a bold attempt at resolution of this antinomy:

> There could conceivably be cases, however, where the dynamical covering law which features in dynamical explanation is itself emergent in the sense that it cannot be derived from the laws of physics. (Walmsley, 2003, p.12)

In such cases, he says, it would be "the law itself which emerged" (ibid). This is an interesting position but there is surely something obscure in the idea of an emergent law. Moreover, even if it is a workable idea it does not seem to mesh well with our usual conception of what emergence should be about – i.e. the phenomenon itself.

Emergence, then, is difficult to pin down. It is in this respect that the highlighting of a parallel with dialectical theory might prove useful. In dynamical accounts emergence is often associated with the idea of "self-organising systems" (Kelso, 1995; Thelen & Smith, 1996). Such a system is:

> ..one in which some kind of higher level pattern emerges from the interaction of multiple simple components without the benefit of a leader. (Clark, 1997, p.73)

We will bypass for the time being the question of whether 'self-organization' and 'emergence' are synonyms, and concentrate instead on the idea that the components are "multiple" and "simple". It is here, we might suggest, that a comparison with dialectics first suggests itself, for what is characterized as 'emergence' in this instance seems remarkably similar to what we have earlier seen described as "the transformation of quantity into quality" (Engels, 1883, p.26). In both cases that which emerges does so on the basis of changes in a quantitatively specifiable substrate but exhibits qualitative properties which are not fully describable merely by reference to these changes. In support of such a parallel we might perhaps note that Kelso and Hegel favour similar illustrations of their respective principles:

> The temperature of water is, in the first place, a point of no consequence in respect of its liquidity: still, with the increase or diminution of the temperature of the liquid water,

there comes a point where this state of cohesion suffers a qualitative change and the water is converted into steam or ice.   (Hegel, 1817a. p.158/9)

Take a little cooking oil, put it in a pan, and heat it from below. If the temperature difference between the top and bottom of the oil layer is small there will be no large-scale motion of the liquid. Notice the liquid contains *very many molecules,* and the heat is dissipated among them as a random micromotion…..This is an open system, activated by the application of a temperature gradient that drives the motion. As this driving influence increases an *instability* occurs. The liquid begins to move as a coordinated whole, no longer randomly but in an orderly rolling motion.   (Kelso, 1995, p.6/7)

Provisionally, then, we might suggest that emergence can be usefully understood dialectically as qualitative change which arises on the basis of changes to what we have called a 'quantitatively specifiable' substrate. In the above examples the latter can be identified with the relatively homogeneous mass of molecules composing the water or cooking oil.

It will be remembered, however, that we have already made a link between dialectics and emergence, for we saw in chapter 3 how Levins & Lewontin had developed their own account of emergence based on consideration of dialectical principles. In drawing attention to this account we highlight a potential problem, for, as we saw, Levins & Lewontin's version of emergence was precisely one which did *not* posit the homogeneity of component parts (L&L, 1985, p.272). Likewise we might also note that in more recent dynamical accounts homogeneity is not always thought to be a requisite. Thelen and Smith (1996), for instance, argue that "the central tenet of dynamic systems is that order, discontinuities and new forms emerge precisely from the complex interactions of many *heterogeneous* forces" (p.37, my emphasis). We might thus question whether our analysis of emergence in terms of the "transformation of quantity into quality" is always an appropriate one. If quantity is linked to homogeneity then the absence of the latter might suggest the absence of the former.

In the case of Lewontin we suggested that his preference for characterizing constituent parts as heterogeneous failed to account for the relativity of the notions of 'homogeneity' and 'heterogeneity'. However, whilst there is some (dialectical) truth in this it does not provide us with a satisfying solution to our current problem, for we were happy to use a one-sided notion of 'homogeneity' when we thought it would serve our

purpose of identifying emergence with the shift from quantity to quality. A better solution is perhaps derived from the consideration that the shift from quantity to quality need not only be predicated on the homogeneity *of constituent parts*. In order to arrive at something that looks like the transformation of quantity into quality, we can just as easily focus on the quantitative changes in a heterogeneous substrate – in which case 'homogeneity' would still reside in the discrete regularity of changes in control parameters. Thus we might say that whilst the applicability of the notion of the 'transformation of quantity into quality' is most clearly illustrated in instances where both conditions apply – where the components are homogeneous enough to be susceptible to quantification and those components are subject to quantitative change (as in Hegel and Kelso's examples above) – this is not a necessity.

Viewing emergence as qualitative change which occurs on the basis of quantitative change perhaps gives us an additional means of characterizing what has proved to be a definitionally elusive (Clark, 1997, p.106) phenomenon. Whilst we are not proposing that it will be applicable in all circumstances it at least avoids some of the problems outlined in Clark and Walmsley's accounts above. It is not particularly reliant on subjective psychological evaluation, does not entail the exclusion of standard examples of emergence and does not involve the reclassification of the type of phenomena which can be said to be emergent.

Reciprocal causation

Consideration of the nature of emergent phenomena leads us once more to the concept of 'reciprocal' or 'circular' causation. We have seen how this concept has been central to dialectical thought since its first articulation in Hegel's Logic (1817a), reappearing in the works of Marx, Engels, Merleau-Ponty and Lewontin. In the literature on dynamics it figures prominently in two areas. In discussions of emergence (e.g. Kelso, 1995, Clark, 1997) it is used to describe the bidirectional influence that substrate and emergent have on each other, whilst on some other accounts (e.g. van Gelder, 1995, 1996) it used to characterize the interaction of coupled systems, both of which are specifiable at the same level of description.

As regards the former, Kelso refers to the reciprocal relation which holds between collective variables (or 'order parameters') – those variables which track development in higher level emergent processes – and changes in the substrate. In the case of boiling oil:

> ..the order parameter is created by the cooperation of the individual parts of the system, here the fluid molecules. Conversely it governs or constrains the behaviour of the individual parts. This is a strange kind of circular causality (which is the chicken and which is the egg?) but we will see that it is typical of all self-organizing systems. (Kelso, 1995, p.9)

Here we find reference to a number of themes that we have already encountered. In the notion that the order parameter "governs or constrains the behaviour of the individual parts" – sometimes referred to as 'downward causation' (Campbell, 1974) – we find an active equivalent of Lewontin's holistic principle that "the properties [of the parts] come into existence in the interaction that makes the whole" (Lewontin, 1985, p.273). Likewise in the question concerning temporal priority at origination ("which is the chicken and which is the egg") we find a return to a dialectical issue already explored in relation to Arthur (1970) and Merleau-Ponty (1945).

We might also note how, like Engels (1883) and Merleau-Ponty (1947), contemporary commentators on dynamics have been keen to differentiate cases of genuine circularity, as above, from mere 'mechanical' bi-directional cause and effect. Clark for instance contrasts reciprocal causal explanation with "catch and toss explanation" (Clark, 1997, p.105). The latter, whilst paying lip service to the bi-directional nature of causal influence, can still be broken down into discrete episodes of unidirectional causation. Thus, for example, in the case of a relationship between an agent and an environment, a clear separation is posited between processes that input to the brain and processes that output to the world:

> The world tosses inputs to the brain, which catches them and tosses actions back. The actions may alter or simplify subsequent computations, by causing the world to toss back more easily usable inputs and so on. (Clark, 1997, p.106)

Both Clark and Kelso argue that this – feedback oriented - approach is inadequate for many types of real world interaction, and their grounds for saying so hinge on recognition of a further shift from quantity to quality. When interactions between

systems (or processes) become too numerous or too complex (in a quantitatively specifiable way) the utility of linear causal explanation simply breaks down (Kelso, 1995, p.9).

Clark's analysis of the way in which accounts of cognition rooted in continuous reciprocal causation (which he abbreviates to CRC) cause "problems" for "standard cognitive scientific models of analysis and understanding" (Clark, 1998b, p.6) is also noteworthy for its distinct similarity to dialectical analyses. It will be remembered that Hegel's critique of the Understanding was that, in contrast to a holistic, reciprocal, process-oriented approach, it stuck to the "fixity of characters and their distinctness from one another" (Hegel, 1817a, p.113). This critique reappeared in Engels' own attack on the 'metaphysical' method of non-dialectical science, which, he argued, observed things "in isolation, apart from their connection with the vast whole…as constants not as essentially variables" (Engels, 1887, p.34). It was also evident in Levins and Lewontin's description of "the alienated world" of "Cartesian reductionism" in which "parts are separate from wholes and reified as things in themselves, causes separate from effects, subjects from objects" (L&L, 1985, p.270). Likewise Clark notes that systems which are considered from a CRC perspective "are unusually resistant to the kind of divide and conquer approach taken by computational cognitive science" (Clark, 1998b, p.6) because the latter "proceeds by isolating a number of distinct sub-mechanisms, assigning them specific roles and plotting the chains of causal influence that flow between them" (ibid). Clark's analysis of CRC here was made with one particular dynamical account in mind, that of van Gelder (van Gelder & Port, 1995; van Gelder, 1995, 1996), and it is to this that we now turn.

van Gelder

Van Gelder describes the following problem faced by eighteenth century industrialists (van Gelder, 1996, pp. 422-424). A steam powered flywheel is used to drive machinery such as a weaving loom. It is important that the speed of the flywheel remains relatively constant. However changes in steam pressure or workload can affect the speed of the flywheel. To keep this speed constant the amount of steam entering the pistons can be

controlled by a throttle valve. The problem is how to ensure that appropriate adjustments are made to the throttle valve as and when necessary. In theory this task could be carried out by a human mechanic but that would be an expensive and inefficient solution.

Two other solutions are conceivable says van Gelder The first is to design a device which measures the speed of the flywheel, compares it with the desired speed and, if necessary adjusts the throttle valve accordingly. This, says van Gelder, is the "computational" solution (ibid. p.426). It involves the utilisation of representations (e.g. the speed of the flywheel) and the making of calculations (e.g. comparing the actual speed of the flywheel with the desired speed) These operations form discrete subtasks – they can be conceptualized in separation from each other and may even be devolved to different components of the device. Moreover the subtasks would be carried out in a sequence, a sequence which would repeat itself in a 'cyclical' fashion.

A solution such as this would not have been possible in the eighteenth century, and more importantly, says van Gelder, would not have been as "direct and elegant" (ibid. p.424) as the solution eventually arrived at. This solution – the 'Watt Governor' - can be described as follows:

A vertical spindle is attached to the flywheel. The spindle has two arms on hinges and on the end of each arm is a metal ball. As the flywheel spins the spindle also rotates, driving the arms outwards and upwards by centrifugal force; the faster the flywheel spins the higher the arms are raised. This arm motion is linked directly the to the throttle valve, affecting the extent to which it opens or closes. Thus if the speed of the flywheel increases, the arms will raise and close the throttle valve a corresponding amount, restricting the flow of steam and bringing the speed back down again. Likewise if the speed of the flywheel decreases below a certain point the arms will come down, thereby opening the valve, increasing the flow of steam and bringing the speed up to the desired point again.

This latter approach to the governing problem is a 'dynamic' approach and it stands counterposed to the stepwise, algorithmic computational approach. There is no compartmentalization of discrete subtasks carried out in a linear and cyclical fashion.

Nor are there any representations and so its "processing cannot be a matter of the rule-governed manipulation of symbolic representations" (ibid. p.427). Instead there is a mutual dependency between the activity of the arms and that of the throttle valve which is ongoing and fluid – each is "simultaneously determining the shape of each other's changes" (ibid).

At this point we might pause to note that van Gelder's account is already interesting to us, if only because it takes as its central metaphor one of the most significant artifacts from the industrial revolution. In this respect it would perhaps not be out of place in Engels' *Dialectics of Nature* – and indeed would outshine many of the examples found there in its detail and clarity. Engels does in fact make reference in this work to the steam engine as "the instrument which more than any other was to revolutionise social conditions throughout the world" (Engels, 1883, p.294) and alludes to significant improvements made by Watt (ibid. p.81). There is also, as we have seen, some suggestion of embodied cognitive developmental interplay between artifact and emerging human powers:

> The hand alone would never have achieved the steam engine if the brain of man had not attained a correlative development with it, and parallel to it, and partly owing to it. (ibid. p.18)

Engels, however, fails too see any instantiation of his own 'dialectical laws' in the operation of the steam engine or any of its components. It is difficult to say whether this might indicate a failure of imagination or whether he took the "mechanical motion" of such artifacts to be at odds with the dialectical principles which were to be reserved for human and natural processes only. The latter position would seem difficult to sustain given that he sought to uncover dialectical laws, such as the 'negation of the negation', in mathematical operations (Engels, 1887, p.188). There is perhaps the additional possibility that as the most prominent symbol, and practical source, of factory mechanization (Marx, 1867, pp.496-499) he would have thought it perverse to link the steam engine with dialectics.

Returning to van Gelder, however, it is, not merely the use of a significant industrial artifact that makes his account of interest to us, it is the fact that he believes that "people bear deeper similarities" (van Gelder, 1996, p.439) to this artifact than to its

computational equivalent. We have already touched on some of his reasons for thinking so in earlier sections. For example the Watt Governor is continuous rather than discrete: "its entire operation is smooth and continuous, there is no possibility of dividing its changes over time into distinct manipulations" (ibid. p.438). It is also centrally incorporates reciprocal causal relations. We should notice here, however, that unlike Kelso's convection rolls example, reciprocal causation is not characterized in terms of 'downward causation'. Rather it is a more democratic "coupling" that obtains between the two main components of the system. There is a dependency between "coupled variables" which "simultaneously, interdependently co-evolve" (ibid. p.437) - and which van Gelder thinks can be applied in a variety of contexts to human systems. Of particular interest is how it is held to apply to agent-environment interaction. Cognitive processes are not considered as one component of the interchange, but rather are implicated in the overall process:

> In this vision, the cognitive system is not just the encapsulated brain, since the nervous system, body and environment are all constantly changing and simultaneously influencing each other, the true cognitive system is a single unified system embracing all three. (van Gelder, 1995, p.185)

We can see how this description recalls the Marxian account of human development arising from the reciprocal interaction between productive agent and environment. The co evolution of coupled variables maps onto the co evolution of agent and environment which is at the root of the Marxian account of man's developing nature. The notion of a unity between agent and environment mediated by physical activity recalls both the process of externalization we outlined in Hegel and Marx's account, as well as the supersession of alienated labour as presented in *On James Mill*. In this respect, as we shall see in the next chapter, it shares some common ground with enactivist accounts, which have consciously incorporated dynamicist elements (Thompson, 2007, pp.38-43). There is however a way of interpreting the notion of an 'extended cognitive system' too literally – a topic we will return to in chapter 7.

Van Gelder's analysis also brings with it other elements which mirror to some extent features of Hegelian/Marxian perspectives. That he is concerned with the agent/environment interaction considered as a totality means that his account, like dialectical accounts – and for essentially the same reasons – can be considered

'holistic'. This holism manifests itself most clearly in the notion of "total state" (van Gelder & Port, 1995, pp.14-15) explanations. As Clark summarises it: "the dynamicist chooses to focus on changes in total state over time" (Clark, 1998b, p.11). Thus for the dynamicist, as for Hegel, "the true is the whole" (Hegel, 1807, p.11) and the methodology can thus likewise be articulated as an attempt "to trace out the internal connection that makes a continuous whole of … movement and development" (Engels, 1887, p.37).

As with Hegel too, an emphasis on holism, continuity and reciprocity, gives rise to a scepticism concerning the usefulness of a static symbolic medium. As we saw in chapter 1, one of Hegel's objections to "logical mechanization" (Hanna, 1996, p.270) was that its representational elements were "fixed in isolation" (ibid) and so inadequate to the task of capturing the world in its fluidity and interrelatedness. A similar position is adopted by van Gelder, for whom the continuous nature of reciprocal coupling renders inappropriate cognitive architectures founded on the manipulation of "tokens interpretable as symbolic representations" (van Gelder, 1996, p.437). As Eliasmith remarks, "'coupling' thus replaces the idea of 'representation passing' for dynamicists" (Eliasmith, 1996, p.445).

We should also note that in talking of "coupled variables" (van Gelder, 1996, p.437) van Gelder is alluding more specifically to the use of differential equations as a tool in the dynamicist armoury for understanding cognitive systems. In this respect his position parallels Levins & Lewontin's suggestion that an account which "breaks down the alienation between the object-organism and the subject-environment must be written as a pair of *coupled* differential equations in which there is co-evolution of the organism-environment pair" (L&L, 1985, p.105, author's emphasis). We will not make too much here of the fact that L&L's account appears to predate that of van Gelder and other dynamicists in cognitive science (e.g. Thelen & Smith, 1994; Kelso, 1995), for a case could be made that L&L, unlike their successors, are not directly addressing the topic of 'cognition'. Nor are we implying that L&L have had any significant influence on these dynamicists (despite the odd references in relation to other topics, e.g. Thelen & Smith, 1994, p.145) as they have had on some enactivists (see next chapter.) Nevertheless, that L&L arrive separately at the same conclusion as the result of an

explicitly dialectical analysis does drive home the appropriateness of the comparison between dynamics and dialectics.

As for L&L's suggestion that the reciprocity expressed through coupled differential equations is one that tracks the break down of alienation between subject and object, this perhaps raises a question concerning the *senses* in which we might also take dynamics to be promoting an 'unalienated' view. We have the fact that, as with L&L's account, it presents agent/environment interaction as a unity, and so as 'unalienated' in L&L's objective, holistic sense whereby lack of separation equates with lack of alienation (L&L, 1985, p.270). However, insofar as there is a normative dimension of alienation, the question perhaps remains of whether the various parallels between Marxian and dynamic conceptions of the agent have radical implications in a wider sense. There is perhaps no *necessary* route from the 'is' of cognitive models to the 'ought' of social change, but a case can at least be made that the non-algorithmic agent-environment interactions described by dynamicism are not only incompatible with a production-line conception of cognition but with the production-line itself. We have noted how the Watt Governor is not pre-programmed in any meaningful sense, and how its operations are "smooth and continuous" (van Gelder, 1996, p.438). As regards the latter, this smoothness is of course entirely mechanical in nature, but as a metaphor (Eliasmith, 1996, pp.447-449) for the unfolding of intelligent human behaviour it is intended to refer to a non-mechanical human reality.

This reality, we might argue, includes the fact that the "self-creation" (Marx, 1844a, p.101) of the embodied human subject is a fluid process, not one composed of discrete pre-programmed steps. It is thus a reality at odds with the imposed and discrete processes of the production line as dictated by the 'detailed division of labour', just as the facts of embodiment are at odds with the mental/manual division of labour. Whilst the restricted rhythms of the production line might be describable in dynamical terms, in the same way that the repetitive swings of a pendulum can be, we might say that the spirit of van Gelder's account of agent activity, as non-algorithmic and fluid, is more in accord with Marx's conception of man's potential as a freely developing being (Marx & Engels, 1848, p.26). Hence, although there is no deductive necessity in the step from dynamic accounts of cognition to the proposition that man is most himself when

unfettered by the division of labour, it is at least plausible to suggest that accounts such as van Gelder's are more congenial to this conception of things.

Hierarchy, Planning and Self-Organization – A Political Interlude

As an addendum to the above discussion of the socio-political implications of dynamical theory, we might also say a little to unify a number of related themes that we have alluded to in various places on the subjects of planning and emergence.

In our introduction we said that, according to our version of Marxian theory, a society which facilitated the free development of its members must be an "emergent non-hierarchical phenomenon". We can perhaps see how such a conception of social possibilities might find sustenance in accounts of self-organized systems where global properties emerge "without the benefit of a leader" (Clark, 1997, p.73). In this respect, the type of social order envisioned is less like an orchestra under the control of a conductor and more like "playing improvised jazz in a small combo" (as Clark describes continuous reciprocal causation, 1998b, p.2).

However, if such a social order is a genuine possibility then, we would argue, certain facts must follow about the agent. We have already touched on this point in various guises at different junctures in this thesis. For example, in chapter 3 we expressed disagreement with Engels' emphasis on the human agent as a 'planner' (Engels, 1883, p.18), although we noted as significant his apparent derivation from this of the appropriateness of a planned economy. Similarly, in chapter 4 we saw how Boden had drawn attention to GOFAI sounding passages in *Capital* where Marx had differentiated the human agent from other organisms on the basis of his planning capabilities (Boden, 2006, p.1029). There we argued that, although this did represent one strand of Marx's thought, the Marx we were interested in was a 'praxis' oriented Marx. On this interpretation human beings are not first and foremost 'planners' at a cognitive level. Planning does not explain human intelligence but is one activity of many which arises out of the more fundamental human capacity for social praxis.

Underlying the positions taken in these two instances is the idea that, an emergent non-hierarchical social order is only a real possibility if it can be shown to be so for the individual agent. In particular it must be demonstrated that the cognitive architecture of the agent is not such as to preclude intelligent spontaneous activity from the outset. It is our contention that certain threads in current cognitive science go some way towards fostering such an outlook. For all of our criticism of Damasio's 'somatic marker' hypothesis (Damasio, 1994, 2000, 2003), for instance, this analysis does at least consider non-rational grounds for intelligent action. Likewise the move away from an algorithmic conception of cognition in dynamical theory (van Gelder – see above), from an emphasis on planning based architecture in Brooks' account of his artificial 'creatures' (Brooks, 1991b), and from detached planning as a necessary precursor to effective action in Kirsh & Maglio's account of 'epistemic action' (see chapter 8) are all, in different ways, friendly to this perspective.

In certain respects, then, the retreat from alienation also incorporates a 'retreat from planning', and on our interpretation of Marx it is necessary that it does so.

Embodiment and Materialism

In this chapter we have so far looked at various ways in which recent cognitive science has adopted ideas and strategies consonant with a Marxian perspective and with our overall theme of a 'retreat from alienation'. In doing so we noted one instance where cognitive science offered an improvement on Marxian theory - van Gelder's clear and detailed account of the operations of the Watt Governor provided an example of a materialist dialectics which was superior to many of the examples found in Engels' work.

Another instance of cognitive science improving on Marx, or at least offering supplementary material for Marxists to consider, can be found in the hypothesis that concepts are often grounded in the facts of human embodiment. A clear expression of this position can be found in Lakoff and Johnson (1999) where it is argued that "the very structure of reason itself comes from the details of our embodiment…" (ibid. p.4)

According to this version of embodiment theory it is not just that our conceptual apparatus need to be physically instantiated somehow or other but that

> …the very properties of concepts are created as a result of the way that the brain and body are structured. (Lakoff & Johnson, 1999, p.37)

Such a position, we might say is not just consistent with Marxian theory, it also supplements it with an insight unanticipated by Marx or later Marxian theorists. We saw in chapter 2 how Marx, in a general way, related the genesis of ideas to our active material being – "the phantoms formed in the human brain are also, necessarily sublimates of their [i.e. human beings'] material life process" (Marx & Engels, 1970b, p.47). Nevertheless, although Marx recognized that man is an "embodied, living, real, sentient objective being" (Marx, 1844a, p.104) he failed to make any explicit connection between the generation of ideas and the embodied nature of individual human beings. Instead he preferred to source our conceptual framework in the wider arena of material social production (Marx & Engels, 1970b, p.42).

As we noted in chapter 3, this approach was then taken up and developed in a particular direction by Sohn-Rethel (1978). Sohn-Rethel wished to establish, amongst other things, that our abstract concepts were neither passively empirical nor a priori transcendental in origin. Instead they must be "derivates from material being" (Sohn-Rethel, 1978, p.201). On Sohn-Rethel's account this material being was located in the practical act of commodity exchange (we will return to this topic in our next section.)

What Lakoff and Johnson's argument shows us, we might suggest, is that, pace Marx and Sohn-Rethel, 'material being' is not only to be located in our productive or economic activities, but also in the fact of our individual embodied existences. The embodiment thesis can thus be seen as an additional route out of the impasse between empiricism and transcendental idealism. As Lakoff and Johnson themselves put it:

> Reason is not, in any way, a transcendent feature of the universe or of disembodied mind. Instead it is shaped crucially by the peculiarities of our human bodies. (Lakoff & Johnson, 1999, p.4)

At the same time, we might note that it is the separation of Lakoff and Johnson's account from other recent cognitive scientific perspectives which prevents it from

developing into a more comprehensive materialist outlook. Were their hypothesis to be hybridized, for instance, with accounts which extend the agent's material being into the environment (van Gelder & Port, 1995; Clark & Chalmers, 1998) the result might be something approaching Marxian materialism, with ideas supervening on a wider agent/environment 'body'.

Not Just Online

Lastly in this chapter we come not to an example of recent cognitive science mirroring Hegelian-Marxian theory – or providing it with supplementary insights – but of a possible Marxian critique of one assumption frequently made by current cognitive science.

Modern (philosophy of) cognitive science will often make reference to a distinction between 'online' and 'offline' cognition (e.g. Clark & Grush, 1999; Wilson, 2002; Wheeler, 2005). Online cognition (or intelligence) is felt to have something to do with immediate interaction with world, such as to produce "a suite of fluid and flexible real-time adaptive responses to incoming sensory stimuli" (Wheeler, 2005, p.12). Offline cognition, by contrast, is more like Hegel's "pure thought" (Hegel, 1807, p.122) or Marx's "pure theory" (Marx & Engels, 1970b, p.52), divorced from a particular material base. It takes place in separation from direct engagement with the world and may include activities like "planning, remembering, and day dreaming, in contexts not directly relevant to the content of plans, memories or day-dreams" (Wilson, 2002, p.626). It is linked in different accounts to abstract thought (ibid. p.625), "internal representations" (Clark & Grush, 1999, p. 9) and reflection (ibid. p.13), and it is sometimes claimed that it can be identified by the fact that its objects are "distant in time and space" (Wilson, 2002, p.635).

Wheeler notes that the offline-online distinction might pose some difficulties insofar "there will be all sorts of hard-to-settle intermediate cases" (Wheeler, 2005, p.12), but he does not think that this detracts from the firmness of the distinction in the general case. However, it seems possible that Wheeler might be wrong here and that there could be something fundamentally problematic about this dichotomy. Wilson's

suggestion, for example, that offline cognition might include planning, daydreaming and remembering "in contexts that are not directly relevant" (Wilson, 2002, p.626) to those plans, daydreams etc. seems to beg the question concerning what notion of 'relevance' is being used here. The world that an individual inhabits, we might argue, is charged with significance such that any feature of it might bear directly on his/her plans, daydreams and memories in countless ways. If I am daydreaming about Paris it is not necessary that I be in Paris for my environment to bear a significant relation to my daydream. Likewise it is not clear in what sense an object's being "distant in time and space" (Wilson, 2002, p.635) should count as a reliable indicator that the subject is indulging in offline cognition. The very act of asking "how far away does an event have to be, or how long ago does an event have to have occurred, for us to regard it as an object of off-line reasoning" draws our attention to a degree of arbitrariness about the question and to the problematic nature of 'pure presence' in the first place.

We might suggest then that it seems possible that the offline/online distinction is another example of a case where something has been divided "into two mutually exclusive and jointly all-encompassing categories [and] it turns out on further examination that these opposites interpenetrate." (Levins & Lewontin, 1985, p.284). To pursue such an analysis would involve arguing that even clear cut examples of offline cognition could be viewed as online in some respect or according to some description, and vice versa. It would also perhaps necessitate an examination of related dichotomies implicit in the online offline dichotomy (e.g. 'object as perceived' versus 'object as represented'.) For space reasons we will not pursue such an analysis here. Nevertheless we might bear in mind this general idea of a fundamental problem with the dichotomous nature of the online/offline distinction if we now turn to a related proposition that has some currency in modern cognitive science, and which appears to run counter to its unifying trajectory. This is the notion that 'proper' cognition is offline, and thus that embodied, embedded etc. accounts of cognition – which in general are thought to line up with 'online cognition' – have a limited sphere of applicability. We find such a conception of things particularly clearly expressed in Clark & Grush (1999) where it is suggested that "truly cognitive phenomena are those that involve off-line reasoning" (ibid. p.12).

We should note that accounts like Clark & Grush's *are* concerned with establishing an essential relatedness of sorts between the online and the offline insofar as they assert that the latter might have its historical origins in the former. Thus, for instance, Wilson (2002) provides many instances of cases which she thinks illustrate the principle that "offline cognition is body based" (p.632). These include Clark and Grush's own argument that offline cognition is derived from the emulation of online mechanisms, as well as examples where working memory is said to make "off-line use of sensorimotor resources" (p.633, in reference to Baddeley, 1986), or where mental concepts can be shown to be rooted "in sensory and motoric knowledge" (p.634), as in Lakoff and Johnson's work (Lakoff 1988; Lakoff & Johnson, 1999).

However, because Wilson's account, and those she cites, have no appetite for challenging the online/offline dichotomy itself, they also show no predilection for questioning the idea that there are distinctive modes of cognition matched to either side of the divide or that one such mode might be more genuinely cognitive than the other.

It is here that we might suggest Sohn Rethel's (1978) ideas could have something to offer, if only in its suggestion of a possibility not hitherto discussed in cognitive science. It will be remembered how Sohn-Rethel sought to derive our propensity for conceptual abstraction from the 'material being' of commodity exchange. Insofar as his argument is only that "commodity exchange is an original source of abstraction" (Sohn-Rethel, 1978, p.28) it could be interpreted as meaning something not too dissimilar to Clark & Grush, or Lakoff & Johnson's, position i.e. that offline abilities have their origin in online activity. However, although these parallels are warranted, as we have seen in chapter 3, Sohn-Rethel goes a little further than this. With the concept of "real abstraction" (ibid. p.19) he does not just want to argue that our capacities for abstraction are derived from practical activity, but that the practical activity is itself an example of abstraction. He is explicitly opposed to the idea that "abstraction is the inherent activity and the exclusive privilege of thought" (ibid. p.18) and suggests instead that, in commodity exchange, "the action alone is abstract" (ibid. p.28).

It is true that Sohn-Rethel's account is more suggestive than clear in its implications. Moreover we would at least need to go into greater detail about the mechanics of exchange and the origins of exchange value before we could hope to get a better idea of

what 'abstract action' might consist in. Nevertheless, we might say that in the very act of positing abstraction as a feature of activity Sohn-Rethel has thrown the cat amongst the pigeons. Abstraction, as a species of cognition, is typically considered to be located firmly within the 'offline' camp; or if, as Wheeler (2005) seems to suggest, the online/offline dichotomy is more like a continuum, then it would be found at the extreme limit of the offline end. If abstraction can be convincingly shown to be a species of action also, this would perhaps suggest the need to reconsider that scheme of things – of a need to reconsider traditionalist assumptions about the cognitive division of labour which still persist in some areas of contemporary cognitive science.

Conclusion

In this chapter our concern has been with the retreat from alienation in cognitive science. We have said that the introduction of embodied, embedded etc. conceptions of the agent constitutes an attempt to 'unify' the agent in a sense comparable with Marx's vision of an unalienated agent – "the whole man" (Marx, 1844a, p.91). This striving for unification utilizes an anti-dichotomous stratagem and in some cases (e.g. dynamics) incorporates a more comprehensive dialectical analysis. In certain instances, such as Noë's praxis-like conception of enactive perception, there is also evidence of Marxian influence (in this case, via Merleau-Ponty).

In addition to outlining a retreat from alienation we have looked at some related issues. We have suggested that, on our interpretation of Marx, the retreat from alienation must be – and is – also a retreat from planning; that Lakoff & Johnson's version of embodiment presents supplementary insights for a Marxian theory of materialism; and that Sohn-Rethel's notion of 'real abstraction' perhaps points to a way out of a residual cognitivism founded on the idea of the inviolability of the online/offline dichotomy.

In our remaining three chapters we will look in more detail at particular areas in cognitive science – enactivism in chapter 6, the extended mind in chapter 7 and epistemic vs pragmatic action in chapter 8. In each case we find further evidence of our hypothesized retreat, as well as elements that are open to criticism from the standpoint of Hegelian-Marxian theory.

Chapter 6 – Enactivism and The Embodied Mind

Introduction

We now turn to an examination of aspects of 'enactivist' perspectives, considered in relation to our theme of the 'retreat from alienation'. This chapter is divided into two halves. In the first half we will consider 'enactivism and autopoiesis' in a general way. In the second we will concentrate on Varela, Thompson & Rosch's 1991 book - *The Embodied Mind*.

Enactivism & Autopoiesis

As suggested in chapter 5, and as many have noted (Di Paolo et al, 2010; Torrance & Froese, 2011), the term 'enactivism' is open to a number of different interpretations. Previously we have used it to characterize Noë's work. In this chapter we use it to refer to a cluster of "intertwined"(Di Paolo et al, 2010) ideas which are influenced in part by the autopoietic theories of Maturana and Varela (1980, 1987) as well as by Varela's later – less explicitly autopoietic – work (1991). Current proponents of this type of 'enactivism' include Thompson (2004; 2007), Torrance (Torrance, 2005; Torrance & Froese, 2011) Di Paolo (Di Paolo, 2005; Di Paolo et al, 2010) and Froese (2009). It should be made clear from the outset, however, that insofar as the following involves allusion to a generic 'enactivist' outlook this is necessarily something of a simplification, concealing differences of emphasis (e.g. between 'enactivism' and 'autopoietic theory' per se), divergent areas of research, and changes in perspective over time.

It is not our intention to look in detail, or systematically, at all of the ideas and their interrelations which make up an enactive perspective. Nevertheless, for the record it is perhaps worth noting that there is general consensus (Torrance, 2005; Thompson, 2007; Di Paolo et al, 2010; Torrance & Froese, 2011) that there are five main themes, even if there is less consensus regarding how these five themes are to be carved up. One

succinct summary of such themes is provided by Torrance & Froese (2011). On an enactivist analysis, to be a (cognizing, conscious) agent is to be:

> a) biologically autonomous (autopoietic) organism – a precarious far-from-equilibrium, self-maintaining dynamic system; b) with a nervous system that works as an organizationally closed network, whose function is to generate significance or meaning rather than…to act via a set of continually updated internal representations of the external world; c) the agent's sense-making arises in virtue of its dynamic sensorimotor coupling with its environment, such that d) a world of significances is 'enacted' or 'brought forth' by a process whereby the enacted world and the organism mutually codetermine each other; and e) the experiential awareness of that organism arises from its lived embodiment in the world. (ibid., p.22)

We perhaps notice a certain imbalance here since a) appears to incorporate 'autopoiesis' per se, and thus implies a number of subsidiary theses regarding what it is to be a living organism (or 'unity'), whereas c) and d) could be viewed as variations on the single theme of agent/environment coupling. Our concern in this section is chiefly with certain of the subsidiary theses associated with a), as well as with c) and d). (We are less interested in b) and e).) These themes are of interest because they appear consonant with ideas that we have already found in Marxian theory.

Unity, Becoming and the Continuity between Life and Mind

Before developing this point, however, we would do well to observe in a more general way how, as was the case with dynamics, enactivist theory often utilizes Hegelian and Marxian sounding ideas and terminology. Perhaps most significant of these is the aforementioned notion of a 'unity', a term used to characterize systems such as a "cell, immune network, nervous system, insect colony or animal" (Thompson, 2007, p.65) in accordance with certain of its features. We will look at one of these features shortly. Here we will just note how the enactivist use of this term compares with Marx and Hegel's own uses of the term. Conceptually speaking a 'unity' for the Hegel of the *Logic* was a dialectical resolution of an antinomy – thus 'reciprocity' was the 'unity' of 'cause' and 'effect' (Hegel, 1817a, p.218). Materially speaking a 'unity' was also what was achieved, according to Marx, when the agent's relationship to the world, other agents and its own activity was no longer fragmented by an imposed division of labour. Perhaps noteworthy in the current context is the fact that, as we saw in chapter 2, this state of affairs was sometimes expressed in biological terms as "the unity of living and

active humanity with the natural inorganic conditions of their metabolic exchange with nature" (Marx, 1855, p.489), a fact which led Fromm to declare that for Marx "labour is man's effort to regulate his metabolism with nature" (Fromm, 1966, p.13).

For enactivists a 'unity' is similarly more likely to be a physically instantiated, biologically specified system, although, like its Hegelian forebear, a certain amount of conceptual dialectics is also involved in its characterization. Thus Maturana and Varela (1987):

> A unity (entity, object) is brought forth by an act of distinction. Conversely each time we refer to an entity in our descriptions, we are implying the operation of distinction that defines it and makes it possible. (p.40)

Here the authors appear to be making the point that establishing the identity of a system, whether that establishing is done in practice by the system itself, or in theory by an external observer, also entails establishing a difference (with other components of the world). This perhaps provides us with an alternative interpretation of Engels' observation that "every organic being is every moment the same and not the same" (Engels, 1887, p.35), for on an autopoietic reading this would not just imply that organisms are in constant flux, but also the stronger thesis that it is only by practically positing its un-sameness with the rest of the world that an organism can establish itself as a 'unity'.

The notion of a 'unity', then, is both important for enactivism and central to the theme of this thesis - for it is precisely a 'unified' account of the human agent which we are arguing constitutes an area of significant commonalty between Hegelian-Marxian theory and recent philosophy of cognitive science. This is not to say that the mere fact that enactivist theory utilizes the concept of a 'unity' implies that it thereby subscribes to every facet of our hypothesized 'retreat from alienation'. Nevertheless, insofar as 'conservation of a unity' might include, on an enactivist account, the conservation of a set of dialectical relations between individual agents and between those agents and their environment, it certainly has some strong affinities with a Marxian outlook. We will return to this point below.

Other Hegelian-Marxian sounding concepts often used in the enactivist camp include that of the 'continuity between life and mind' (e.g. Thompson, 2004, Di Paolo, 2005)

and 'becoming'. Whilst the former might not seem an obvious candidate for Marxian comparisons we might say that, as with the Marx of *The German Ideology*, it suggests a general predilection for starting from the "material life process" (Marx & Engels, 1970b, p.47) and deriving consciousness from this. In this respect Marx's assertion that "Life is not determined by consciousness, but consciousness by life" (ibid) seems particularly amenable to an enactivist interpretation.

As regards 'becoming' (e.g. "we are constituted in language in a continuous becoming" – Maturana & Varela, 1987, p.235) we will not look at this notion in detail here. This is partly because we will return to it briefly later. It is also because there is perhaps only a limited amount of ground to be gained from pursuing comparisons where the parallels are chiefly with the 'Hegelian' side of an 'Hegelian-Marxian' outlook. It is not news to current proponents of enactivism that portions of their approach are heavily indebted to the phenomenological tradition. Indeed this fact is everywhere acknowledged and explored – even if the names more commonly mentioned are Husserl, Heidegger and Merleau-Ponty, with the founder of phenomenology, Hegel, generally neglected. However some parallels are more significant than others and relate as much to Marxian theory as to Hegelian phenomenology. It is to these that we now turn.

Self - Production

One striking and central affinity between enactivism and the Marxian perspective lies in the autopoietic conception of living beings as 'self-producing'. Thus Maturana and Varela (1987) assert that:

> …Living beings are characterized in that, literally, they are continually self-producing. We indicate this process when we call the organization that defines them an autopoietic organization. (p.43)

and this view of things is restated in more recent enactivist works, with Froese & Di Paolo (2009), for example, alluding to the centrality of "material self-production" (p.443) to autopoiesis. As we have seen in chapter 2, there does not seem to be much which separates this conception of living beings from Marx's own conception of the defining characteristic of human agency. Man, as 'Homo Faber', was precisely "the result of his own labour" (Marx, 1844a, p.82) and so can likewise be understood in

terms of a 'fundamental circularity' – "by thus acting on the external world and changing it, he at the same time changes his own nature" (1867, p.283). It was this idea of the "self-creation of man" (Marx, 1844a, p.101) that Marx had taken originally from Hegel and then concretized by annexing it to the human material life process rather than to the self-development of an abstract 'spirit' or 'idea'.

Of course, in drawing attention to this parallel between autopoietic and Marxian theory we are also perhaps drawing attention to a divergence. Self production, on the autopoietic account, is characteristic not just of human beings but of all living beings – and perhaps more generally of 'systems' of a certain sort (Maturana & Varela, 1980; Thompson, 2007; Froese & Di Paolo, 2009). Thus in one sense the autopoietic account could be seen as undermining the Hegelian-Marxian account which posits self-production through labour as a defining feature of the human agent. To the extent that this is true we might take this as another area where modern cognitive science has improved upon Marxian theory. Marx, we might say, had rather narrowly imposed a strict line of separation between the human mode of 'becoming' and the mode of becoming of other organisms. Autopoietic theory has therefore established the existence of continuities where Marx had only seen a difference.

However we perhaps should not make too much of this point, for it might be argued that the Marxian account of human self-production does contain elements which are absent from the autopoietic account, and these are perhaps sufficient to differentiate human self creation from mere organismic self creation. In particular the self-production of the human agent, on a Marxian analysis, is developmental and open ended. Through labour human beings develop over time and, considered as a species, their potential to become 'fully human' – to be able to "produce in a human manner" (Marx, 1844b, p.121) – is contingent upon further events in the ongoing history of their self-creation. Moreover the portion of the spatial environment by means of which they realize themselves is also not a fixed quantity but an ever expanding one (e.g. through agriculture). This would seem to contrast with the situation of a simple autopoietic entity such as a cell. Whilst the latter "emerges as a figure out of a chemical background" (Thompson, 2007, p.99) its ongoing self-productive activity serves only to keep it in a stable state. Its "circular process of self-generation" (ibid) is not open-ended

or developmental in the same manner as human self-production, and its continued existence is dependent on the maintenance of spatial boundaries.

Agent and Environment

In drawing attention to the self-production of the organism we also draw attention to other features of an enactivist outlook. Indeed it is this fact itself, as much as the nature of such features, which suggests Marxian parallels, for 'self-production' is not a standalone component of enactivism. Rather, as on a Marxian account, it is internally related to other components. Di Paolo remarks in this respect that "these internal relations bespeak the strength of the associations under a single banner" (Di Paolo et al., 2010, p.4). We might say that our thesis of a 'retreat from alienation' is one that is in broad agreement with such a principle – although it would like to increase the number of features between which internal relations are held to obtain and would also like to write something else on the banner.

Both enactivism and Marxian theory, then, start from the fact of 'self-production' and in elaborating on this elicit other internally related features. As one would expect if such internal relations are genuine, there is also a degree of agreement between Marx and enactivism on what some of these other features might be. Thus, for example, both seem to agree that self-production is also production of an environment. We have already examined in some detail the different senses in which this is true on a Marxian account and will not catalogue these senses again here. However by way of a brief summary we might say that, according to Marx, Man 'realizes' himself through productive interaction with the material world and this interaction is described in terms of a reciprocal relationship between agent and environment. We might also remind ourselves here that, contra Hegel, Marx thought there was a necessary separation of sorts between agent and environment but nevertheless that in dialectical fashion this separation also implied a unity: "A being that does not have its nature outside of itself is not a natural being" (Marx, 1844a, p.104). In unalienated labour this unity manifests itself in the fact that the agent's product becomes an expression and extension of his being, so that there is no longer felt to be any separation between producer and product.

There are similar elements in the enactivist account. There too we find talk of self-realization. Thus for example Maturana and Varela assert that through their autopoietic organization organisms "become real and specify themselves" (1987, p.48) and, in phraseology similar to Levins & Lewontin (1985), this self-realization is likewise presented as arising from a reciprocal coupling between agent and environment where the "enacted world and the organism mutually codetermine each other" (Torrance & Froese, 2011, p.22). Moreover, as with the Marxian account such codetermination is double-edged. Whilst, as we have seen, a unity is said to bring itself into being by distinguishing itself from the environment it is also true of living beings that "their organization is such that their only product is themselves, with no separation between producer and product" (Maturana & Varela, 1987, pp. 48-9), thus in this sense the modus operandi of the autopoietic entity resembles that of the unalienated labourer. However we perhaps ought to avoid over simplification here. Whereas on a Marxian account it is the self-same external environment that an agent is separate from and stands in reciprocal relation to, on an enactivist account things seem a little more complicated. On a narrowly autopoietic perspective the non-separation between producer and product seems to apply only to the internal milieu of a "self-constructing *closed* system" (Maturana &Varela, 1980, p.v – my emphasis). Thus, strictly speaking its product is not a construct out of external resources in the same sense as that of the Marxian agent. On the other hand, a broader, phenomenologically oriented, enactivist picture does seems to be one which incorporates the wider world into the package – thus Thompson: "the external world is constituted as such for the system by virtue of the system's self-organizing activity" (2007, p.27.)

In remarking on this close relationship between agent and environment in enactivist literature we should also note how such a theme is developed not only in relation to its implications for the self-productive subject but also in relation to its implications for our understanding of the object. In particular we should note the centrality for enactivism of the idea that the world of the agent is "brought forth or enacted" (Thompson, 2007, p.13). Our use of the phrase 'the world of the agent' here is deliberately vague for that which is brought forth is variously characterized as "a world or cognitive domain" (Di Paolo et al, 2010, p.5), "a world of significances" (Torrance & Froese, 2011, p.22), "the external world" (Thompson, 2007, p.27), and "an environment" (ibid. p.59). We will not explore such differences here but instead will

just note in general fashion how this 'bringing forth of a world' also has Marxian parallels. We saw in chapter 2 how one version of 'praxis' likewise entailed that 'the world is brought forth through practical activity'. Some commentators, as we noted, had interpreted this in an overly metaphysical fashion, suggesting that on Marx's account "reality is a product of the dialectical interaction of nature and human nature" (Rubinstein, 1981, p.170). However, we argued that a non-metaphysical reading was possible which did not call into question the objectivity or otherwise of the material constituents of the world. It seems that similarly non-metaphysical readings might apply to enactivist accounts which ground themselves in the objectivity of biological systems. As we shall see in the second half of this chapter when we look at Varela, Thompson & Rosch's 'Embodied Mind', one such reading, influenced to some degree by Levins & Lewontin, uses the notions of the enacting or bringing forth of the world as a counter to conceptions which view the agent as a passive victim of a pre-given environment to which he must adapt

A Dialectical Whole

We have said so far that enactivism has similarities with Marxian theory insofar as it is concerned with the material self-production of the organism and links this self-production with production of an environment. We might add that it does not just concern itself with the relationship of the organism to itself and to the environment but also, on some accounts (e.g. Maturana & Varela, 1987; Thompson, 2001; Torrance & Froese, 2011), with the relationship of the organism to other organisms. We have already given some examples of this in the case of human agents when we broached the topic of intersubjectivity in the previous chapter. Here we will just note how, with this third element in place, the enactivist outlook can be seen as presenting a 'dialectical whole'. This is particularly well illustrated in a diagram produced by Maturana & Varela (1987) a version of which is given below:
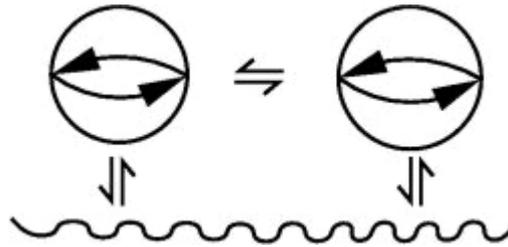
'Third-Order Coupling' (from Maturana & Varela, 1987, p.180)

This diagram shows the "reciprocal structural coupling" (p.180) of two organisms (possessing nervous systems) with each other and with the environment. It can be viewed as 'dialectical' at least insofar as it is both holistic – with each component implicated in the activity of the others – and (as with the dialectics of Hegel, Marx, Engels, Merleau-Ponty and Lewontin) incorporates circular causal relationships.

In terms of the 'retreat from alienation' we might say that the state of affairs depicted by the diagram has great similarities with that found in Marx's account of free labour (e.g. in Marx, 1844b). To make this comparison clear we might modify the diagram as follows:
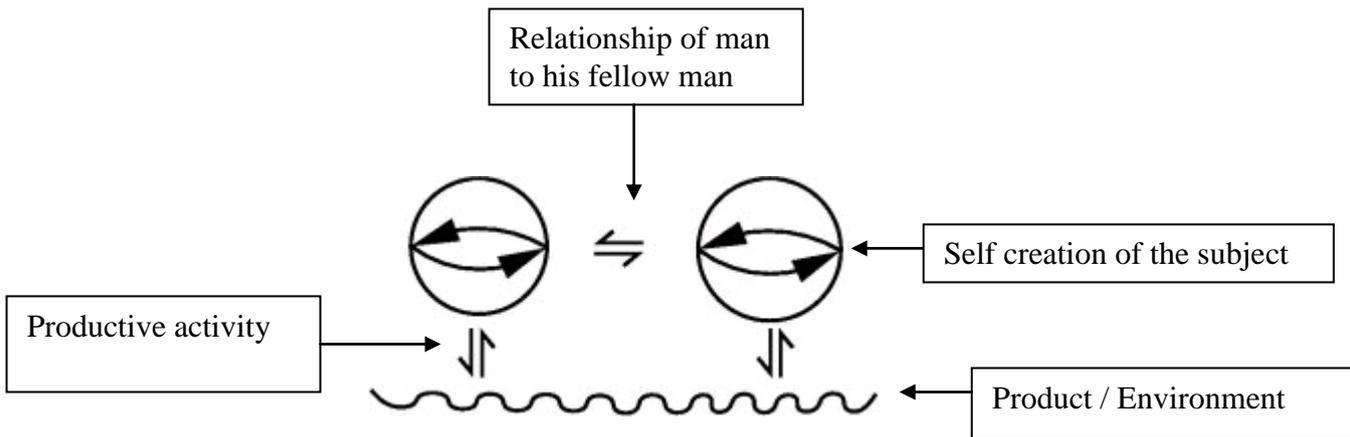
"Our products would be so many mirrors in which we saw reflected our essential nature"

(Marx, 1844b, pp. 121-122)

The labels used in the modified diagram, we would argue, do not deviate greatly from the original intentions of Maturana and Varela, and perhaps just represent a 'special class' in this respect. We should note that although each component has been given a separate label, because of the internal relation between component processes, it would also be possible to use each label to describe the whole. Thus, for example, on a Marxian analysis the whole process could be viewed as 'the self-creation of the subject' for the latter arises from activity that is both social and productive. Similarly 'the relationship of man to his fellow man' is mediated by his productive activity and the results of this productive activity. (This latter point is illustrated by the quotation from *On James Mill* given at the bottom of the diagram.) Likewise again, strictly speaking the product and environment of product activity includes other agents as well as oneself – a point which is partially acknowledged by Maturana and Varela (1987, p.180). We should also note that the same diagram (perhaps with fuzzy arrows!) could also be used to illustrate the situation of the alienated labourer (as described in Marx, 1844a) who finds himself alienated from his product, his activity and his fellow men.

A Dialectical Problem

In saying above that the autopoietic/enactivist conception of a dialectical whole mirrors the Marxian account of free labour we are edging our way closer to a thesis that the enactivist outlook is in some way favourable to the normative ideal of freedom found in Marxian theory. We will develop this point shortly. Before doing so however we might note one respect in which an autopoietic outlook, or at least one argument for it, is perhaps less than dialectical. This argument is not central to our concerns in this section but is worth remarking upon in passing. The argument in question is given in Maturana & Varela (1987):

> Reproduction cannot be a part of the organization of living beings because to reproduce something, that something must first constitute a unity and have an organization that defines it. This is simple logic… (p.57)

It is perhaps surprising to find this line of reasoning in a work which in other respects has elevated circularity to a principle of life itself. Our dialectical objections to it are

twofold. The first echoes an earlier objection made to Arthur (chapter 2). It is not clear why, given that the authors accept the inherent circularity of self-production, they should find less acceptable any circularity inherent in positing reproduction as a defining feature of living systems. That they do view the latter as problematic perhaps suggests that they have some residual deference for 'chicken/egg' style reasoning (Kelso, 1995, p.9; see also Varela et al, 1991, p.172) - in this case almost literally so.

Secondly, given that one establishes the temporal priority of constitution to reproduction this does not seem to give one grounds for excluding the latter as part of the organization of a living being considered as a unity. To characterize reproduction as exterior to the organization of the system in this way seems to be to adopt an unwarrantedly fragmented approach. It would seem to involve a temporal prejudice according to which that which occurs after a certain defining event is viewed as inessential (at least to an account of what constitutes a living system.) Such a prejudice is perhaps incompatible with a process-oriented approach. In its place we might recall Hegel's 'temporal holism' which viewed living systems in terms of "a necessary and complete process of becoming" (Hegel, 1807, p.20) rather than as fixed quantities comprehensible at a particular instant in time.

It should be noted, however, that in directing this criticism at one twenty five year old autopoietic text, the intention is not to imply that such criticism applies across the board to present day enactivism or autopoietic theory. Insofar as our argument is that Maturana & Varela failed to give 'temporality' due consideration, for example, we might note that Di Paolo proposes an updated autopoiesis which, amongst other things, gives increased prominence to temporality (Di Paolo, 2005, pp.13-14).

Alienated Labour and Autopoiesis

We have found, then, many affinities between an enactivist perspective and a Marxian outlook – these include emphasis on the material self production of the agent, agent-environment codetermination and the conception of these processes (expanded to include inter-agent interaction) as a dialectical whole.

These affinities may not be entirely coincidental. As we have remarked already, Varela's (1991) work, to which we shall turn shortly, was influenced to some degree by the dialectical outlook of Richard Lewontin. Likewise, as briefly alluded to in chapter 3, the work of the enactivist Evan Thompson (2007) draws heavily on that of the Marxian phenomenologist Maurice Merleau-Ponty, and particularly on the dialectical approach taken in *The Structure of Behaviour* (Merleau-Ponty, 1945). We might add also that, according to Maturana at least, the notion of 'autopoiesis' grew out of the revolutionary ferment of May 1968, and in particular in the occupation of the University of Chile, of which Maturana was a part (Maturana & Varela, 1980, p.xvi). We can perhaps surmise from this that Maturana was likely to have become acquainted with some Marxian ideas during this period.

Such affinities as we have described however, have mainly been of a 'theoretical' sort. They have not implied any necessary common ground on political, social or normative issues. Whilst there is some talk in praise of 'revolution' and against 'reform' in recent enactivist accounts (e.g. Di-Paolo et al, 2010) such phrases are not to be understood as having any application outside the arena of philosophic paradigms. This is in keeping with the majority of the parallels between a Marxian outlook and the unifying trajectory of modern cognitive science described in this thesis. Although we have made links between particular conceptions of cognition and particular modes of social organisation, we have not suggested that such links are explored by the relevant authors themselves. The retreat from an alienated conception of the agent in cognitive science, insofar it remains restricted to the confines of cognitive science, has no need to align itself with any particular political or social program.

Nevertheless, we might conclude this half of the chapter by noting that, in the case of enactivism, this line *is* occasionally crossed. One significant example is Maturana's introduction to *Autopoiesis and Cognition*. Here Maturana appears to argue that for human beings to realize themselves as autopoietic systems a certain type of social organisation is necessary (Maturana & Varela, 1980, p. xxv). In such a system the activities of human beings "as social organisms must satisfy their autopoiesis" (ibid. p.xxvi). This move, we might observe, represents a transition from a description of the structural pre-requisites of living systems to a recommendation for a particular type of social order. However, rather like Marx's human 'essence' (Marx, 1844a, p.80) or

'species being' the transition is made on the basis of what is necessary for the proper realization of the organism. There is thus a sense in which it can still be taken as non-normative on the grounds that it is simply describing what conditions must prevail in order for a particular organism to function as it 'should'.

In elaborating on this position Maturana alludes to a social state of affairs wherein human beings are unable to satisfy their autopoiesis:

> A human being that through his interactions with other human beings participates in interactions proper to the social system in a manner that does not involve his autopoiesis as a constitutive feature of it, is being used by the social system but is not one of its members. (ibid. p. xxix)

Here comparisons with the Marxian account of alienated labour seem unavoidable. In both cases the object of criticism is a social state of affairs wherein human beings participate in a collective activity that does not facilitate the realization of their essential nature, yet where such activity is felt to be 'proper' to the system (an expression of its rationale) and is 'used' by it. Likewise, in both cases, insofar as this social system is not organised in accordance with the essential nature of human beings – their potential to realize themselves freely through activity – those human beings are 'not one of its members' i.e. are objectively 'alienated' from it.

Maturana goes on to specify what kind of social system would facilitate human autopoiesis and comes to the conclusion that "such a society is in its essence an anarchist society" (ibid. p.xxx). Whilst we do not wish to play down the differences between a Marxian outlook and an anarchist one (see e.g. Thomas, 1980) it will be remembered that the version of Marxian theory which informs our thesis is of a 'libertarian' or 'humanist' variety and as such is relatively friendly to anarchism. If Maturana intends by his use of the phrase 'anarchist society' not just a society with an absence of hierarchy (Maturana & Varela, 1980, p.xxix), but one with an absence of coercion in all spheres of life, where production is not imposed on the individual by force or as an economic necessity but arises from his own "spontaneous activity" (Marx, 1844a, p.80) then there is little scope for disagreement here.

The Embodied Mind

We have looked briefly, then, at enactivism and autopoiesis and noted how enactivist accounts contain many elements amenable to a Marxian interpretation. In this second half of the chapter our aim is to look at one particular text from the enactivist camp - Varela, Thompson & Rosch's book 'The Embodied Mind' (1991). This might seem like an odd way of ordering things. However a look at Varela et al's book in separation from a general analysis of enactivism seems warranted. This is so not just because of the influential nature of this text but also because it contains additional distinctive elements relevant to the perspective of this thesis. It should be noted that the following is not intended as a summary of the most significant parts of the text. Indeed much of what is generally taken to be important about the work will be absent from our account, which will focus only on a few areas that are key to our analysis.

The Influence of Lewontin

As mentioned earlier, one respect in which Varela's 1991 book is of particular interest to us is in the reference it makes to Richard Lewontin (Lewontin, 1983; Levins & Lewontin, 1985)[12]. Dennett suggests that Varela et al. "draw heavily on the claims by the geneticist Richard Lewontin that evolution must be understood from an enactive perspective" (Dennett, 1993, p.125). In fact Dennett seems to be conflating two points here, for while Varela et al. draw heavily on Lewontin's work (in one section of the book), Lewontin did not describe his own position as "enactive". Nevertheless the central point remains valid that Varela's enactivism is significantly influenced by Lewontin.

The authors introduce Lewontin in a discussion of what they refer to as 'natural drift'. The latter is a conception of evolution which rejects the notion that selective pressures involve the optimization of fitness traits. The evolutionary process is not to be viewed in a prescriptive fashion as one that "guides and instructs" (Varela et al, 1991, p. 195)

---

[12] We have thus far referred to Levins & Lewontin's (1985) text 'The Dialectical Biologist'. Varela et al's book refers to Lewontin's (1983) paper. However a version of this paper appears as a chapter in the 1985 book attributed to both authors.

but rather in a proscriptive fashion as one that confines its operations to the discarding of traits that are incompatible with survival. It is concerned with 'satisficing' rather than 'optimizing' (ibid. p.196).

For Varela there appears to be a direct connection between this position and Lewontin's own position, insofar as both in different ways are opposed to a conception of the agent as passive product of environmental forces. The move from conceiving of "selective pressures as broad constraints to be satisfied" (ibid. p.198) to the idea that "the very notion of what an environment is cannot be separated from what organisms are and what they do" (ibid) is viewed as one that can be taken in a single step. Thus Varela proceeds to endorse Lewontin's 'anti adaptationist' argument as a logical adjunct of the natural drift argument.

As we have seen (in chapter 3) Lewontin's argument hinges on a dialectic between organism and environment. We find all the elements of this dialectic reproduced in Varela's account. Firstly there is the general notion that "the organism and the environment are not actually separately determined" (ibid p.198, citing Lewontin, 1983). Taken by itself this could perhaps be interpreted as meaning nothing more than that full specification of an environment necessitates specification of its inhabitants and vice versa. There is an internal relation between the two because "just as there is no organism without an environment, so there is no environment without an organism" (ibid).

Full blooded co-determination, however, implies more than this. It includes, for example, the fact that the organism will alter its environment. The environment is not just found but 'enacted'. We have already seen on Lewontin's account how this position implied the organism's involvement in the literal creation of its physical environment. In the case of human agents Varela adds to this a cognitive component such that "the world we cognize is not pre-given but enacted through our history of structural coupling" (ibid. p.200). Clearly, as suggested in our survey of enactivism the term 'world' here is susceptible to multifarious interpretations – relating as much to the world as experienced as to the world objectively constituted in a physical environment (insofar as such a distinction can be made.) We will not explore this issue further here however.

In addition to an analysis of the effect of the agent on the environment a complete account of agent environment co-determination also involves recognition of the fact that the environment has a reciprocal influence on the agent. Thus "organisms have constructed environments that are the conditions for their further evolution" (ibid. p.202, citing Lewontin, 1983). As we saw in chapter 3 in Lewontin's case this full characterization of the agent-environment dialectic is expressly linked to Engels' own account of evolution in *Dialectics of Nature*. Hence we have a clearly traceable path from Engels to one of the key components of Varela's enactivist outlook.

Three Points of Convergence

The Embodied Mind is not just of particular interest to us because of its debt to Levins & Lewontin however. It also noteworthy because of an argumentative thread which runs through the book and which has three significant points of convergence with a Marxian account, above and beyond those which we highlighted in our account of enactivism earlier. We might summarise these points as follows:

1) The fragmentation of the individual subject
2) The social roots of this fragmentation
3) An explicitly pragmatic (rather than 'theoretical') solution to this fragmentation.

We will deal with each of these in turn.

1) The fragmented subject

Varela et al. state in their introduction that:

> The existential concern that animates our entire discussion in this book results from the tangible demonstration within cognitive science that the self or cognizing subject is fundamentally fragmented, divided or non-unified. (Varela et. al., 1991, p. XVII)

and this concern with the fragmentation of the subject is indeed reiterated throughout the book. Here then, as in Marx's *Economic and Philosophic Manuscripts*, the central problematic is given as that of the disunified subject.

With respect to the characterization of this disunity, the parallel with Marx is not an exact one, but there is significant overlap. For Marx, as we have seen, the disunity of the alienated labourer consisted in her separation from other individuals, from her own activity (and hence her alienation from self) and from the product of that activity (Marx, 1844a, p.77-96). In relation to the mental/manual division of labour this fragmentation was presented in terms of the severance of bodily being from mental activity (Marx & Engels, 1970b; Sohn-Rethel, 1978).

The disunity of concern to Varela et al. is certainly bound up with the some of these. They are to some degree concerned with inter-agent unity (see next section) although they do not explore this issue in any detail. More notably they are concerned with a number of fragmentary and antinomous conceptions of the subject all ultimately related, on their account, to the philosophical mind-body problem (see next section). These include, as with Noë's later account (and likewise influenced by Merleau-Ponty), the separation of perception from action, for which a similarly praxis-like remedy is proffered - "studying cognition not as recovery or projection but as embodied action" (Varela et. al., 1991 p.172). They also include the separation imposed by traditional cognitive science between experience and cognition (a theme also taken up in later enactivist accounts e.g. Varela & Shear, 1999; Thompson, 2004, Thompson, 2007).

This latter, we might note, appears to have some internal relation to a wider critique of the dichotomy between "science" and "experience". In this respect the authors see themselves as following in the footsteps of Merleau-Ponty in their search for an "entre-deux" between these two poles. In their journey they confront this opposition in different guises according to context – as 'realism' vs 'subjectivism' (Varela et. al., 1991, p172), as 'objectivism vs subjectivism' (ibid. p. 230) and, rather indirectly, as 'absolutism vs nihilism' (ibid. p.143). Opting for either option in exclusion of the other is seen, in Hegelian fashion, as 'one-sided' (ibid. p.123) and gives rise to an inadequate conception of the predicament of the agent.

In one section of the book, appropriately titled "Selfless Minds: Divided Agents' the situation is described in terms reminiscent of the 'unhappy inwardly divided consciousness' which is aware of the "self-contradictory nature of itself" (Hegel, 1807, p.125). Cognitive science, as represented particularly by Minsky (1986) and Jackendoff (1987), operates by hiving itself off from the realm of experience, says Varela. When it examines the human agent it is unavoidably drawn to the conclusion that the Self (or 'consciousness', or 'free will' - there is some blurring here, Varela et. al., 1991, p.128) is redundant, either as a concept or as an efficacious force. Yet at the same time we as human beings (including human beings who work in Cognitive Science, such as Minsky and Jackendoff) find ourselves unable to abandon these constructs:

> Cognitive Science tells us that we do not have a Self that is efficacious and free. We cannot, however, give up such a belief – we are "virtually forced" [as Minsky says] to maintain it. (Varela et. al., 1991, p.129)

It is noteworthy that when this dichotomy manifests itself as the separation between experience and sub-personal processing, the authors berate traditional cognitive science for dismissing or ignoring the former, and thereby presenting a "radical and alienating view of unconscious processing" (ibid. p.48). Taking their terminology at face value we might suggest that they have here highlighted another sense in which traditional cognitive science could be viewed as 'alienated'. In terms of the arguments we gave in chapter 4 this disregarding of conscious experience, or its reduction to an epiphenomenon (Thompson, 2007, p.13), might perhaps also be viewed as a result of a 'production line' conception of cognition. An account which characterizes agent mentality in terms of inputs, outputs and intermediate processing is free to dispense with consciousness as a causally irrelevant by-product, like smoke from a factory chimney.

For purposes of our current argument, however, it is less important that we pay heed to the, sometimes confusing, detail of Varela et al's version of the fragmented subject than to the fact that they take the fragmented subject as their starting point.

2) The social basis of fragmentation

Like Marx then, Varela et al. start out from the fact of the fragmented subject. Having made the case for fragmentation the authors go on to relate the theoretical expression of this fragmentation to social practices which are also fragmented. Thus in the case of the 'mind body' problem:

> It is because reflection in our culture has been severed from its bodily life that the mind-body problem has become a central topic for abstract reflection.
> (Varela et. al., 1991, p.30)

Here we find, in condensed form, several of the theses already explored in our account of Marx in chapter 2, and in particular the Marx of the *German Ideology*. At the most general level we might say that the statement indicates a recognition that our conceptual outlook and concerns have their roots in social practice as "sublimates" of the "material life process" (Marx & Engels, 1970b, p.47). As such it is also consonant with that materialist approach which "does not explain practice from the idea but explains the formation of ideas from practice" (ibid. p.58). In Varela's case it is not Marx but Lakoff and Johnson (see chap.5) who supply the materialist backdrop here, with the argument that our conceptual propensities can be traced back to "the structured nature of bodily and social experience" (Varela et. al., 1991 p.178, citing Lakoff, 1988), but the outcome is the same.

As well as revealing a generally materialist orientation, however, the above passage from Varela et al explicitly mirrors elements of the Marxian critique of the mental/manual division of labour. As we have emphasized, Marx's critique of capitalist social relations is in part concerned with the enforced separation of mental and manual activity brought about by the mental/manual division of labour. This "separation of head and hand" (Sohn-Rethel, 1978) at the material level is not only thought reprehensible for the material fragmentation it imposes on the agent but is also held responsible for a variety of aberrant productions at the conceptual level. These include a conception of the agent as fundamentally fragmented. The "organic and social unity" (Pannekoek, 1937, p.448) of mental and physical activity is neglected in our characterizations of the agent because "the division of labour separated these two parts into functions of different classes" (ibid). Varela et al can thus be seen as making an

analogous connection when they link the theoretical "mind-body problem" with the material separation of thought from practical activity in the wider culture.

In stating that the authors' account mirrors Marx's critique of the mental/manual division of labour, we do not mean to imply that they have made any systematic connections between the economic or productive organization of society and the fragmentation of the individual psyche. In fact such topics do not figure largely in their account. Nevertheless, it should be noted that when capitalist social relations are alluded to, it is in a critical way and one which at least suggests a critique of the fragmentation of the social whole. The authors are for example, against "the economic view of the mind" (Varela et. al., 1991, p.245) where the goal of the self "is assumed to be profit – getting the most at least cost." In contrast with Damasio (1994, see our chap. 4) they will have no truck with a "business deal mentality" (Varela et. al., 1991, p.246) and they see as misguided the "view of the self as an economic man", a view which they think is prevalent in the social sciences. Perhaps confusingly – given the centrality of the idea of 'boundaries' to autopoietic theory – they link such a conception of man with the idea of boundary maintenance:

> The self is seen as a territory with boundaries. The goal of the self is to bring inside the boundaries all of the good things while paying out as few goods as possible. Since goods are scarce, each autonomous self is in competition with other selves to get them. (ibid)

It is in countering such a conception of things that the authors make one of their few allusions to inter-agent unity, for here they wish to emphasize that "self-interest is always other directed" (ibid. p.247). Like the Lord in Hegel's parable (Hegel, 1807), those who seek to attain self-realization through competition with their peers are engaged in a contradictory and ultimately futile enterprise – they are "struggling in a confused way to maintain the sense of a separate self by engaging in self-referential relationships with the other" (Varela et. al., 1991, p.247). The solution to this alienation of man from man is a pragmatic one – 'mindfulness/awareness practice' – which can be used to achieve a social unification of sorts, albeit by getting rid of selves altogether (ibid, pp.59-63).

Coming full circle, it is also mindfulness/awareness practice which, according to the authors, provides a practical solution to the separation of mind and body; and it is to this idea of a practical solution to a theoretical problem that we now turn.

3) A Pragmatic Solution to Fragmentation

Our concern in this section is not with the content or likely efficacy of the activity of "mindfulness/awareness practice". This topic will be broached shortly. Our interest is rather in the mere fact of its selection as a solution to the problem of fragmentation, for it is here that the pragmatic orientation of Varela et al's account can be most clearly seen. From the outset the authors make known their pragmatic intent. In the introduction they state:

> ..those human traditions that have focused on the analysis, understanding, and *possibilities for transformation of ordinary life* need to be presented in a context that makes them available to science. (ibid. p. xv – my emphasis)

In some respects this program is not too dissimilar to our own. Marxian theory (and practice) might also be regarded as one of those 'human traditions' which aim at understanding and transforming ordinary life – and we also are attempting to present it in a context which makes it available to cognitive science. The parallel we want to emphasize, however, is not this meta parallel (between Varela's selection of meditative traditions and our focus on Marxian theory) but rather that between Varela's own conception of the pragmatic application of 'mindfulness/awareness practice' and Marx's conception of a practical solution to theoretical problems. As we have seen (in chapter 2) it is part of the Marxian outlook that merely theoretical solutions to philosophical problems can fail even on their own terms. The eleventh 'Thesis on Feuerbach' which states that "the philosophers have merely interpreted the world, in various ways, the point is to *change* it" (Marx 1970a, p.123) carries with it not just a normative call to action but the suggestion (if not a logically deducible one) that in some instances the world can only be adequately understood through action. A similar meaning can also be attributed to Marx's remark that "all mysteries which lead theory to mysticism find their rational solution in human practice and in comprehension of this practice" (ibid. p.122). As regards this latter statement, however, it will be remembered that were several possible interpretations. On one account it is the understanding of

human practice which facilitates solutions to philosophical problems. We saw in chapter 5 how this might apply to Noë's account of enactive perception. It was a comprehension of the active nature of perception which enabled him to counter some of the assumptions made by the 'argument from analogy' and the 'grand illusion' arguments. On a stronger interpretation, however, the above statement can be taken as suggesting that it is engagement in action itself which can facilitate solutions to philosophical problems; and on the strongest interpretation it is the action which *constitutes* the solution to a philosophical problem. It is this last interpretation which is implicit in Marx's declaration that "the solution of theoretical oppositions is only possible in a practical way" (1844a, p.93) and in Fromm's statement that "the resolution of *theoretical* contradictions is possible only through *practical* means" (Fromm, 1966, p.29).

Varela's enactivism seems to mirror Marx's position in this respect – for Varela too seems to endorse a strong kind of pragmatism whereby theoretical problems are given a practical solution. We note for instance that Varela et al want to differentiate their position from passive phenomenological approaches, and in so doing dispense with the 'merely theoretical'. Husserl, say the authors, was unable to bridge the gap between science and experience because his approach "completely lacked any pragmatic dimension" (Varela et. al., 1991, p.19). Likewise Heidegger and Merleau-Ponty[13] "both stressed the pragmatic, embodied content of human experience, but in a purely theoretical way" (ibid).

Not confining oneself to the "purely theoretical" in Varela et al's case means recognizing that the mind-body relation is not a fixed quantity but is something that can be changed. Descartes' view of himself as a 'thinking thing' was a product of his original question ("whether body and mind are one or two distinct substances" - ibid. p.28) which in turn was "a product of specific practices – those of disembodied, unmindful reflection" (ibid) A different view is attainable, one involving recognition that "body and mind can be brought together" (ibid), if one adopts reflective practices that themselves bring about this unification:

---

[13] Here, of course, Varela et al are not referring to the Merleau-Ponty of *Humanism & Terror* or *Adventures of the Dialectic*.

> The basic assertion of the progressive approach to human experience is that the mind-body relation or modality is not simply fixed and given but can be fundamentally changed. (Varela et. al., 1991, p.28)

From a certain (analytical) perspective this style of argumentation can seem flawed or incoherent. We want to know if mind and body are irreconcilably separate things. Varela's response is 'they can become one'. This seems not to answer the question without further elaboration. Either the authors mean that mind and body were always one (in some metaphysically definite sense) and it is just that a particular style of reflectiveness can make this unity more apparent to us. Or they mean that mind and body might indeed be irreconcilably distinct substances (or properties or …) but that a subjective, experiential sense of unity can be achieved through the adoption of a particular reflective practice. In either case, it might be argued, the practice itself is not factually relevant to the original question about the relation between mind and body – for there must be an answer to that question which is true irrespective of any particular reflective practice.

Varela et al's point, however, seems to be that this is not the case, or - less radically – that the line between theoretical and practical solutions is not as distinct as, 'Western' philosophy might suppose. The practice (of mindfulness-awareness meditation) is crucially relevant to the question not just as a possible aid to its solution but in some sense as constituting part of that solution. One of the consequences of a "pragmatic approach to the transformation of experience in everyday life" (ibid. p.244) is that the question of the relation between mind and body, when posed from the standpoint of inactive theoretical philosophy, comes to be seen as empty and abstract. A merely theoretical answer to such a question stands the risk of reinforcing that abstractness and the disembodied practices that underpin it. The call to meditation is therefore simultaneously a pointer to a means of finding a solution to the fragmentation of the subject and a recommendation for the abandonment of theoretical practices that help to maintain this fragmentation.

Limitations of The Embodied Mind

We have suggested, then, that *The Embodied Mind* has significant commonalties with a Marxian approach. Insofar as its starting point is the fragmentation of the individual subject, it recognizes the practical/social roots of this fragmentation and proposes a pragmatic solution, the parallel with Marx is clear. Moreover insofar as its enactivist outlook is influenced by Lewontin's work it has a direct line of descendancy from Marxian thought. To this extent we can say that it constitutes a good early exemplar of the 'retreat from alienation' in cognitive science. Nevertheless the book is clearly not – by any stretch of the imagination – Marxian in its outlook; and while we do not think that a Marxian perspective is a pre-requisite for truth in cognitive science, in this case it does seem that the author's account might benefit from being more Marxian in certain respects.

Perhaps surprisingly the limitations of Varela's book have been expressed in a way agreeable to the general orientation of this thesis by Daniel Dennett. Dennett, somewhat satirically notes that "Reform, as we know, is the enemy of revolution" and goes on to affect disappointment at not finding anything revolutionary in the book, concluding that "I think I can proceed with business-as-usual cognitive science" (Dennett, 1993, p.122). One of the problems, he suggests, is that Varela et al are too accommodating to the views of traditional cognitive scientists such as Minsky. "Do these authors despise no one?" (ibid. p.124) he asks ironically, and observes that "This kinder, gentler vision of cognitive science runs the risk of diluting its revolutionary impact" (ibid).

Dennett's tone, of course, is only half serious and his use of 'revolutionary' talk is intended only for application within the realms of theoretical cognitive science. However we might take seriously the general point that Varela et al's position is in certain respects too 'liberal' to be revolutionary on any terrain.

At the most abstract level this is perhaps signalled in their approach to dialectics. There is much that might count as dialectical in their work. We have seen how the authors take up Lewontin's materialist dialectic in describing the relationship between organism and environment. Likewise their central enactivist hypothesis is one that posits a unity between action and cognition/perception (Varela et. al., 1991, pp.172-

173). Moreover there are many passages that we haven't addressed, where the authors explicitly incorporate dialectical insights from Eastern philosophy. These include endorsement of the 'Madhyamika dialectic' (ibid. p.228) as well as the philosophy of Nagarjuna. According to the latter "causes and effects, things and their attributes, and the very mind of the inquiring subject and the objects of minds are equally codependent on the other" (ibid. p.224), ideas that would be at home in Hegel's Logic, even if they predate that work by 2000 years or so.

Nevertheless when it comes to confronting the overarching dichotomy between science and experience, the authors approach seems less thoroughgoing. We have already said both that they are against 'one-sided' solutions to dichotomies, and that they advocate a practical solution to the problems posed by the separation of mind and body. However, in the case of science and experience, the outcome aimed at does not seem to be a synthesis of any kind but rather a "middle way" (ibid. p.230). Thus, for example, in a passage ripe with socio-political overtones, they remark that:

> Unless we move beyond these oppositions the rift between science and experience in our society will deepen. Neither extreme is workable for a pluralistic society that must embrace both science and the actuality of human experience.   (ibid. p.13)

Here, the talk is ostensibly of 'moving beyond oppositions', yet the tone adopted is one of reconciliation rather than synthesis or supersession. The 'pluralism' alluded to appears to be one which recognizes the positive aspects of both antagonistic outlooks and is content to let them co-exist - an interpretation also borne out by a later remark that "experience and scientific understanding are like two legs without which we cannot walk" (ibid. p.14). The point is not to supersede or unify the two elements but merely to recognize their complementarity. With our Hegelian hats on, then, we might say that the "middle way" does not so much suggest the forging of a new unity which transcends the categories of 'science' and 'experience' considered in separation. Rather it suggests the fostering of an agreement between the two spheres that leaves both of them extant as autonomous entities. Thus Dennett would seem to be right that this liberal solution leaves him free to pursue business as usual.

Moreover this "failure of spirit" (Thelen & Smith, 1994, p.41)[14] in the dialectical realm seems to be matched by a failure of spirit in the authors' conception of practice. We have highlighted parallels between Marx's concept of a practical answer to a theoretical question and Varela's call to (meditative) practice. 'Mindfulness-awareness practice' is Varela's pragmatic answer to the problem of the fragmented subject. However we should perhaps question whether it is up to the task.

It will be recalled (chapter 2) how the Marxian critique of the bondsman's supersession of alienation through simple perseverance was that "only the bondsman's consciousness had been transformed" (Rees, 1988, p.39). Those material factors which had been the original source of the bondsman's alienation from his activity and the world had not themselves been the subject of any transformation, and thus it was difficult to see how the bondsman could pull off this trick. A similar critique might be directed at Varela et al's supposition that the fragmentation of the individual could be overcome through mindfulness awareness meditation. The authors describe the unification achieved by this technique as follows:

> When the mindfulness meditator finally begins to let go rather than to struggle to achieve some particular state of activity, then body and mind are found to be naturally coordinated and embodied. (ibid. p.29)

The problem here would similarly seem to be that only the meditator's consciousness has been transformed. There has been a "revolution in thought" (Rees, 1988, p.39) but none in material relations. We are not proposing, of course that the authors have erred in not seeing the necessity for social revolution. Rather we are suggesting that, on their own terms, if there is a problem with the fragmentation of the subject, and if this fragmentation is susceptible to a practical solution, then it seems logical that this must be one which embraces their own enactive conception of the agent. Their critique of traditional cognitive science is one which seeks to replace the separation of thought and activity with "cognition as enaction" (Varela et al, 1991, p.206). This crucially implies the non passive, practical engagement of the agent in different activities. Likewise their insistence that "mind and world arise together in enaction" (ibid. p.177) and that the environment "is enacted by histories of coupling" (ibid. p.204), as we saw earlier,

---

[14] Thelen & Smith characterize connectionism's insufficient radicalism in these terms, cited also by Eliasmith, 1996.

entails that this practical engagement is with a material world that extends beyond the individual subject's body. Like Marx's 'whole man', then, insofar as Varela et al's agent is a unified being she is an active, embodied being, embedded in an environment.

Yet meditation implies the precise opposite of this. It is an inward looking, passive, solitary activity where the individual detaches herself from interchange with the material world. Like the activity of Hegel's stoic it involves "withdrawing from the bustle of existence into the simple essentiality of thought" (Hegel, 1807, p.121). It is therefore difficult to see how it could be the means of unification for their fragmented subject.

Conclusion

In this chapter, then, have we looked at enactivist accounts in general and concluded that they incorporate a number of Marxian themes. These include (but are not exhausted by) the self-production of the agent, and a dialectical conception of the relationship between individual organisms, and between those organisms and the environment, as mediated by activity, which replicates Marx's own account of free labour. We have also looked at *The Embodied Mind*, noting, amongst other things, the influence of Lewontin, as well as a significant argumentative thread which starts off from the fact of the disunified subject, locates its roots in social practice and seeks to resolve it by practical means. Whilst this is also a Marxian strategy, the suggested practical solution (meditation) seems flawed and at odds with an active, embodied, embedded conception of the agent.

Chapter 7 – The Extended Mind

A being that does not have its nature outside of itself is not a natural being. (Marx, 1844a, p.104)

Introduction

In the latter half of the previous chapter we looked at a single work which has had significant influence on current cognitive science – Varela et al's *The Embodied Mind*. We argued that this text, whilst forming part of our hypothesized 'retreat from alienation' also had some weaknesses on its own terms, and that these were weaknesses that a Marxian analysis could help to highlight. In this chapter we turn from the embodied mind to the extended mind. A similar story can be told here, although we ought to note two complicating factors. Firstly, although in this case there is again a main work which will be the focus of our attention – in this instance a single paper (Clark & Chalmers, 1998/1998a, reprinted in Clark, 2008) – there are also other texts by one of the original authors (in particular, Clark, 1997; 2003a; 2003b; 2008) – which to varying degrees deal with the same topic. Secondly, and connectedly, there are not only different texts by the same author but also seemingly different senses of the 'extended mind' given not only between texts but also within texts. Our task will therefore also involve giving some indication of which versions of the 'extended mind' we think are friendly to the 'retreat from alienation' and which present difficulties. In fact, a great deal more time will be spent on the latter rather than the former – although this should not be taken as indicating a devaluing of the former.

We will begin with a summary of Clark and Chalmers original (1998) paper:

The Extended Mind

Clark and Chalmers' (henceforth C&C) initial target in this paper is the idea that cognitive processes are to be located only "in the head" (C&C, 1988, p.8). In place of such a conception of things they wish to advocate what they call an 'active externalism'. To illustrate what is meant by this they point out that some of the tasks which can be completed using cognitive processes (traditionally so called) can also be completed using a mixture of cognitive processes and manipulation of the external environment. Taking examples from the work of Kirsh and Maglio (Kirsh & Maglio 1994; Kirsh 1995 – see next chapter) they argue for instance that one can assess the 'fit' of 2 dimensional shapes into 2 dimensional sockets in a game of Tetris either by imagining them rotated in various ways or by actually rotating them using a rotate button. Similarly, when playing Scrabble, one can either 'mentally' rearrange letters to create a new word or one can physically rearrange the tiles. Moreover, in many cases taking the latter course of action makes the task easier and can in addition be said to be 'enabling' in a variety of ways (see Clark 1997, 2003a, 2008 for development of this theme)

C&C's next move is to suggest that in those instances where we do take the physical manipulation route rather than the imagination route, there is no reason why our doing so should be seen as choosing the 'non–cognitive' option. The justification for this position is what Clark later (2003b) comes to describe as a 'parity principle':

> If as we confront some task, a part of the world functions as a process which, were it done in the head, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world is (so we claim) part of the cognitive process. (C&C, 1998, p.8)

The traditional conception of the cognitive, then, is wrong – cognitive processes "ain't (all) in the head" (ibid). With the case of the scrabble players for instance we can say that "In a very real sense, the re-arrangement of tiles on the tray is not part of action; it is part of thought" (ibid. p.10)

Having made the case for extended cognition C&C then want to "take things a step further" (ibid. p.12). They want to show that not only cognitive processes but also mental states are susceptible to extension. In particular they want to show that "beliefs can be constituted partly by features of the environment" (ibid). To this end they devise the now well-known scenario involving 'Otto' and 'Inga'. Otto has Alzheimer's disease, thus impairing his memory. In order to help him remember things he writes down any new information he 'learns' in a notebook which he carries with him. Included amongst his notes are addresses of buildings. Unlike Otto, Inga has normal memory function. She and Otto hear separately of an exhibition at the museum and both decide to go. Inga pauses to recall where the museum is located (53rd Street) before walking there. Otto looks up the address in his notebook before walking there. In both cases we can say, according to C&C, that the individual concerned wanted to go to the museum and believed that the museum was on 53rd Street. We can say this because "the notebook plays for Otto the same role that memory plays for Inga" (ibid. p.13). Moreover, in both cases it is legitimate to characterize the individuals concerned as having had their beliefs before they consulted their respective information sources; we can say of Inga that "the belief was sitting somewhere in memory waiting to be accessed" (C&C, 1998a, p.7) and there is no reason why we cannot say the same of Otto with his notebook. Hence we have shown that in the case of Otto "a belief is simply not in the head" (C&C, 1998, p.14)

C&C are keen, however, not to open the floodgates to any kind of external information bearing artefact. They thus outline further conditions (later modified in Clark, 2003b) that the notebook fulfills (and which any other contending artefact or medium ought to fulfill) in order to count as part Otto's extended mind. These are that:

1) The item in question must be a constant in Otto's life.

2) The information it contains must be directly available without difficulty.

3) Otto must automatically endorse this information.

4) Such information must also have been consciously endorsed previously.

Having developed this notion of extended belief, C&C conclude matters by indulging in some speculation about further possible applications of their thesis. "Socially

extended cognition" (C&C, 1998, p.17) is one such possibility. If having a belief can be partly constituted by having a notebook it could also be partly constituted by having a human being who is as readily accessible and reliable as a notebook. Thus "the waiter at my favourite restaurant might act as a repository of my beliefs about my favourite meals…In other cases one's beliefs might naturally be seen to be embodied in one's secretary, one's accountant or one's collaborator" (ibid. pp.17-18).

Another such possibility is "the extended self" (ibid. p.18). Our mental states (such as belief) make up a large part of who we are, say C&C, so if we can talk about these being extended it would seem to follow that we can talk about the self being extended. Two possible consequences of this are that "interfering with someone's environment will have the same moral significance as interfering with their person" and that "certain forms of social activity might be reconceived as less akin to communication and action and more akin to thought" (ibid).

We should also note that the argument is advanced in several places by the introduction of 'cyborgian' style speculation about the physical integration of human beings with technical hardware (a theme Clark develops in greater detail in 2003a). Thus, for example, a third 'Tetris' scenario is considered wherein an individual has a neural implant that allows him to rotate the shapes just by willing it. This for C&C "seems to be a case of cognitive processing" (C&C, 1998a, p.2). Likewise, in response to an argument that 'coupled systems' (as exemplified by Otto and his notebook) are too easily decoupled to count as cognitive, C&C imagine a future scenario wherein an individual could plug 'modules' into his brain to help with various tasks such as geometric reasoning. In such a situation, they say, we would not deny that the processes involving these modules were "part of thought" (C&C, 1998a, p.5) despite their being easy to decouple. And again, when attempting to counter the objection that Otto's notebook lookups have a perceptual phenomenology and so cannot be treated as true cases of 'belief', C&C make reference to a scenario in the movie 'Terminator'. The eponymous hero of that film has memories which are "displayed" before him in his visual field, but these displayed items, say C&C, still count as examples of "standing memories" and hence of "standing beliefs" (C&C, 1998, p.16).

The Parity Argument

An initial critical response to this argument might be one that focuses on the 'parity principle'. This is not an untypical move. As Menary points out "the parity principle has become the focus for some of the main criticisms levelled at EM" (Menary, 2010, p.5); and it does seem that there is perhaps a certain sleight of hand involved here. We might highlight this in our own way by drawing comparison with a passage from Marx on the subject of money:

> I may have no intellect, but money is the true mind of all things and so how should its possessor have no intellect? Moreover he can buy himself intellectuals and is not the man who has power over intellectuals not more intellectual than they? I who can get with money everything that the human heart longs for, do I not possess all human capacities? Does not my money thus change all my incapacities into their opposite? (Marx, 1844a, p. 109)

There are perhaps two arguments at work here. One, implicit in the assertion that "money is the mind of all things", recalls Marx's account of commodity fetishism (see chapter 2) according to which we really do, collectively, externalize ourselves in money and commodities, which as a result appear to be endowed with minds of their own (Marx, 1867, pp.163-179). We will return to this idea later.

The main argument, however, is an exploration of the hypothesis that an individual acquires certain capacities as result of acquiring money. Here then, we might say that Marx is putting forward his own version of a 'parity principle' where a different artefact - money - takes the place of the notebook. Despite my lacking certain capacities, money enables me to get results which are identical to those that I would get if I did have those capacities. There is thus a functional equivalence between having those capacities and having money – to such an extent that having the latter can be identified with having the former.

Given that Marx's account also centres on mental capacities, if we were to take it literally we could perhaps construct a thought experiment parallel to that of the Otto story. 'Rich Otto' does not have a notebook but instead has money. He wants to get to

the Museum of Modern Art on 53rd Street. He consults his wallet and pulls out a £20 note. He uses this money to pay for a taxi journey (or a bus ride, or a train journey – the 'vehicle' is unimportant). Could it now be claimed that the money is part of Rich Otto's extended mind? Presumably we would need to show that the money fulfils the same conditions that the notebook is said to fulfil. An attempt at this might run as follows:

1) *The item in question must be a constant in Otto's life*. This is surely the case with the money and so little argument is needed here. A sceptic might suggest that Otto does not have the self-same money available to him at every instant – but even C&C would probably dispute the relevance of this. If the original Otto had a helper at home who periodically transferred the information in the original notebook into a new notebook this would not make any difference to C&C's argument.

2) *The information it contains must be directly available without difficulty*. The issue here is perhaps with what constitutes the 'information' in the case of the money. There seem to be several possibilities. We might take what is literally written on the note as the information (e.g. £20 – or 'I promise to pay the bearer on demand £20 etc.'), or maybe just the visual appearance of the note, or maybe, at a more abstract level, what that note means to Otto in that particular situation (here we are perhaps alluding to an 'affordance' (Gibson, 1979) based approach to information, though re-envisioned for economic life). In the latter case the money might be taken as expressing information along the lines of "I can get you from A to B". In any case, however we interpret the concept of information here, there is no problem with the idea that the money conveys information of some kind.

3) *Otto must automatically endorse this information* – Otto has no reason to question what money tells him, so this is unproblematic.

4) *Such information must also have been consciously endorsed previously* - Likewise Otto has put his trust in money before.

We can imagine many possible objections to this account – objections perhaps centring on the role that transport plays in the 'Rich Otto' story, or on the difference between the types of information presented in both scenarios; or perhaps of relevance here would be Adams & Aizawa's remark that "not just any functional equivalence is cognitive equivalence" (Adams & Aizawa, 2010, p.135). Some of these objections could perhaps be met – although probably at the expense of stretching a definition or two. Our aim, however, is not to develop an elaborate and detailed defence of this story but rather to present it as a kind of 'reductio', for the point with Marx's 'parity principle' is surely that it is not intended to be taken literally. Despite the implication that the moneyed individual has the relevant capacities in a certain sense, Marx's critical intent is to show that the literal construal of this state of affairs is flatly false. The individual who lacks an intellect does not gain genuine intellect just because he has the purchasing power to acquire intellectual results produced by others. A distinction might be made here between social appearance and the 'fact of the matter'. It may be the case that within a particular mode of social organization, having a certain amount of money can have consequences commensurate with having a certain intellect, even perhaps consequences for how the moneyed individual is perceived by others in these terms – but there is also a fact of the matter outside of the arena of social consequence and social agreement. An individual does not gain certain capacities just because the use of a particular artefact in a particular social setting produces results identical to those which would be produced if he did have those capacities.

It is clear how this argument might have implications for Clark and Chalmers' account. At the very least, we might say, they should not take it for granted that functional equivalence (in a particular social setting) entails property equivalence. There is a fact of the matter outside of the social consequences of Otto's use of the notebook. In his case the fact of the matter is that he has Alzheimer's and therefore – on one reading at least - does precisely lack the properties that the Extended Mind argument seeks to attribute to him. No amount of functionally equivalent consequences can change this fact.

Externalization and Extension

An attack on the parity principle, however, is not our chief concern in this chapter. Rather our main interest is in distentangling that which is both valid and chimes with Marxian perspectives, in Clark and Chalmers' account, from that which seems to have gone awry. Our starting point is that of a general sympathy with Clark & Chalmers' position for, unlike non-Marxian critiques of the extended mind, we begin from the assumption that there are senses in which agents *do* externalize themselves in various ways in the material world. Moreover some of these senses appear to match up with versions of the extended mind thesis.

We might remind ourselves in this respect how Marx thought that individuals realized themselves through productive activity and that this process involved a dialectic between subject and object which he expressed as follows:

> By thus acting on the external world and changing it, he [the productive individual] at the same time changes his own nature.   (Marx, 1867, p.455)

We saw in chapter 3 how Vygotsky (1930) made use of this idea in relation to our 'cognitive natures' by introducing the notion of 'psychological tools'. On Vygotsky's account we develop our cognitive capacities through the creation and manipulation of external resources such as signs, diagrams and maps. This strategy "enhances and immediately extends the possibilities of behaviour" (Vygotsky, 1930, p.5). Vygotsky's account, in turn has been very influential on Clark, in particular forming the backdrop for his own reworking of the concept of 'scaffolding' in (Clark, 1997), and it is clear that Otto's notebook in the 1998 paper can be viewed as a psychological tool or 'scaffold' in this sense. Thus we might suggest that there is a 'strand' or .version' of the 'extended mind' thesis, and one to which Clark sometimes confines himself, which makes use of the notion of scaffolding - along with other relevant concepts such as Kirsh & Maglio's 'epistemic action' (see next chapter) - to describe what might be called a 'figurative' extension of mind. Here what is implied is not the literal extension of mind in space, but rather the metaphorical 'extension' of cognitive capacities through the manipulation of external resources.

It should be made clear from the outset, then, that we have no objection to this sense of 'extended mind'. Indeed this is grist to our mill for it is an example of a Marxian idea filtering down to contemporary cognitive science.

Returning to Marx, however, it will also be remembered that his account of productive activity *did* include an element of spatial extension. He was concerned not just with the developmental consequences of man's interaction with nature but with the various ways in which the product of the agent's activity becomes an extension of his being. Nature was the "inorganic body of man" (Marx, 1844a, p.81) not just because it provided man with a means of physical sustenance but also because in altering it man saw himself reflected in it, and in using it, it became a practical extension of him. It was in relation to this latter point that Marx argued that in fashioning tools from natural resources "nature becomes one of the organs of his activity, one that he annexes to his own bodily organs" (Marx, 1865, p.456). We saw in chapter 3 how this idea was taken up by Merleau-Ponty (1945). In his account of finding one's way with a stick, the latter became "a bodily auxiliary, an extension of the bodily synthesis" (ibid. p.176). Again, in turn, this account has had some influence on Clark's own work. In particular, in (2003a, pp.89-114), Clark, referencing Merleau-Ponty, considers the use of artefacts to extend ones *subjective sense* of space. This idea of 'experiential augmentation' seems to give us a second sense of 'extended mind' (although not one that is broached in the 1998 paper.) This version of an 'extended mind' also seems unobjectionable, and is again of interest to us because of its apparently Marxian lineage.

More generally, we have no objection to the idea that agent and environment can be fruitfully thought of as a single system in certain circumstances, or that the interaction between them can be characterized as a single systemic process. As we have said in relation both to dynamical and autopoietic accounts of cognition, this conception of things meshes well with Marx's own view of labour. It is precisely a unity between agents, their activity and their environment which Marxian theory sees as present, in distorted form, in alienated labour and which it thinks can be realized "in a human manner" (Marx, 1844b, p.121) with the end of alienation. Moreover we can agree that 'cognition' is implicated in this reciprocal process and that, to a certain degree, it is a matter of context and convenience whether one attributes it to an individual agent conceived of in separation from her activity, or whether one includes her activity itself

(which inevitably incorporates features of the environment) as part of her cognition. Thus, for instance, there is nothing necessarily problematic in C&C's assertion that the rearranging of scrabble tiles is "part of thought" (1998, p.10) – although it is not clear why, in one-sided fashion, they should therefore conclude that it is "not part of action" (ibid).

All of these examples, then, provide legitimate senses of 'mind extension', senses which have some grounding in Marxian accounts of labour and externalization. We should note however that none of the senses of 'externalization' found in Marx's work imply the *literal*, non subjective, spatial augmentation of statically conceived mental contents or their literal transference to external artefacts. It is this version of the extended mind, as exemplified in C&C's interpretation of the Otto story, which, we might suggest, presents the most difficulties.

Literally extended?

Before pursuing this line of reasoning, however, we should perhaps sound a note of caution, for some may object that we are presenting a 'straw man'. Clark and Chalmers, it might be argued, never explicitly advocate the literal, objective, spatial augmentation of mind. Rather, what they are concerned with are conditions for being "part of the physical substrate of a cognitive system" (Clark, 2008, p.88) or the "local mechanistic supervenience base" (ibid. p.118) for cognition and mental states, and an extended supervenience base is not the same thing as a literally extended mind.

If this objection is both coherent and correct, then all well and good – our critique is not then directed at any version of Clark and Chalmers' extended mind thesis but at a common misconstrual of it, albeit one which is popular (see e.g. Fodor, 2009). It is true, we might observe, that C&C for some reason decided to call their original paper 'the extended mind' rather than the 'extended supervenience base for cognition' but this rhetorical flourish should not be taken too literally.

.We might want to consider, however, how the objection that C&C's real concern is with 'supervenience bases' or 'material substrates' is supposed to play out as a counter

to a literal interpretation of mind extension, particularly in relation to the latter section of the 1998 paper where, as noted earlier, C&C make the transition "from cognition to mind" (C&C, 1988, p.12). In this section C&C unambiguously state that their concern is with "truly mental states" (ibid) which "can be *constituted* partly by features of the environment" (ibid. my emphasis). These mental states are "simply not in the head" (ibid. p.14) not because spatial talk is thought to be misguided but because they are located elsewhere (in the notebook).

At this juncture a critic of the 'literal' interpretation of these passages may point out that C&C's intention is not to suggest that mental states 'themselves' are located in the notebook but merely that the notebook acts as the physical substrate of such mental states. C&C can remain agnostic about whether mental states themselves are located anywhere.

The problem with this line of reasoning seems to be that C&C don't appear to be agnostic about the location of mental states (and cognitive processes) when they aren't being extended. The suggestion rather seems to be that un-extended mind is "in the head" (ibid. pp. 7, 8, 9, 10, 14, 15). Indeed it is this spatial conception of things that justifies talk about 'extension' in the first place. Once we accept that Inga's beliefs are "in her head" the implication is that Otto's are "in his notebook". Otto's beliefs, the argument goes, stand in some relation to his notebook which is equivalent (in relevant respects) to that in which Inga's beliefs stand to her own head/brain. Whatever that relation is (e.g. supervenience), if by virtue of it such phraseology is justified in Inga's case, then by virtue of that same relation such phraseology is also justified in Otto's case. Hence any agnosticism that might be attributed to, or professed by C&C (see e.g. Dartnall, 2004) in relation to the 'actual location' of mental states drops out of the equation as practically irrelevant. C&C's argument boils down to this: "In any sense in which it is meaningful to say that Inga's beliefs are 'in her head' (and C&C do think it is meaningful), it is just as meaningful to say that Otto's are in his notebook."

The Alienated Mind

In making our case against this 'literally extended mind' we might draw further comparison with Hegel and Marx. Earlier we argued that none of the senses of externalization found in Marx's work involved the literal, non-subjective, spatial augmentation of mental contents or their literal transference to external artefacts. What we do find however is the description of scenarios wherein certain artefacts are treated *as if* this literal leap from head to world had been achieved, for this is precisely the case with alienation. As we saw in chapter 1, on Hegel's account of alienation the externalized products of consciousness manifest themselves as "an alien reality already present and given, a reality which already has a being of its own and in which it [consciousness] does not recognize itself" (Hegel, 1807, p.294). Likewise we saw in chapter 2 how Marx took this theme up, characterizing the product of alienated labour as "an alien being, as a power independent of the producer" (Marx, 1844a, p.78) and later extending this analysis also to commodities, which "take on a fantastic form" (Marx, 1865, p.436) appearing as "independent beings endowed with life" (ibid). It is just such an analysis, we might argue, which applies to Clark and Chalmers' literally extended mind. In their interpretation of the Otto story they present us with what can be usefully considered as an 'alienated' account of mind.

In saying that the Otto story presents an 'alienated' account of mind we do not wish to imply, of course, that there is anything amiss in Otto's mode of externalizing himself in his notebook as there is, on a Marxian analysis, with the externalization of human mental properties in commodities for example. In this sense the analogy with the alienated object does not hold up. Otto is perfectly aware that his jottings in the notebook are his own product and he does not relate to them as to an alien being or as to something endowed with its own life and energy. As noted earlier, Otto's activity with the notebook forms a straightforward and unproblematic example of the sort of phenomenon that Clark, under the influence of Vygotsky, links with the term 'scaffolding'.

However the problem occurs when Clark goes the extra mile and describes the information in the notebook as constitutive of Otto's beliefs, or when, as seems to be implied, the notebook itself is seen as constitutive of Otto's memory. It is this

interpretation of the state of affairs with Otto and his notebook which seems to meet the criterion for Hegelian/Marxian alienation. By projecting mental qualities onto the notebook C&C have succeeded in making human mentality itself seem strange and 'other' to the reader. Our everyday acts of externalization now appear as something mysterious to us. Like the commodity which evolves out of its brain "grotesque ideas, far more wonderful than if it were to begin dancing of its own free will" (Marx, 1867, p.164). Otto's notebook seems to have developed ideas above its station.

It is this very strangeness, we might add, which has done so well for Clark and Chalmers, for the Otto story reads like a piece of 'fantastic' fiction - one which flies in the face of "common sense" (Clark, 2008, p.105) - and has thus remained popular partly because it represents what might be called 'a good read'. In this way it far surpasses what Clark himself calls "the single best piece of philosophical fiction ever written" (Clark, 2003a, p.90) – Dennett's short story about spatial displacement 'Where Am I'? (Dennett, 1981a, pp.310-323).

Reification of mind

This 'making strange' of mind occurs partly, we might say, as a result of reification. In projecting mental qualities onto the note book C&C have turned mind, or a subset of mind, into a kind of 'thing' with a specific spatio-temporal location. This reification, in turn, seems to be co-occurent with an objectified account of *internal* mind. Mind, even when it doesn't 'seep out' (Clark, 2003b, p.8), is still thingified on C&C's account.

Such would appear to be the case, for example, with the 'portability criterion' for mind - that to count as part of an individual's cognitive system an artefact needs to be 'reliably available' – meaning in Otto's case that he always carries his notebook with him (Clark, 2003b, p.6). This is thought to be on analogy with Inga's carrying her beliefs around in her head even when she isn't inspecting them. Both cases we might say are founded on a literalistic conception of what it is to 'have a belief' - beliefs are things that you carry around with you, like a piece of personal property, either in your head or in a notebook.

This literal-mindedness is also evident in a later argument put forward by Clark (2003b, 2008) in reply to the objection (attributed to Martin Davies) that the access which Otto has to his notebook is perceptual, unlike the 'access' which Inga has to her memories, and as such is directed at an object to which others also have access. Clark's response (omitting some complicating references to 'multiple personality disorder') is as follows:

> ..Why then suppose that uniqueness of access is anything more than a contingent fact about standard biological recall? If in the future science devised a way for you to occasionally tap into my stored memories, would that make them any less *mine* or part of my cognitive apparatus?...it seems to be at most a contingent fact that I and I alone have a certain kind of access to my own biologically stored memories and beliefs. (Clark, 2008, pp.100 -101, author's emphasis)

There are many assumptions made in this passage which could perhaps be questioned, but we will not look at them all here. Instead we will just make the single observation that Clark has taken as his starting point a traditional 'privileged access' perspective – according to which I have 'access' to my own beliefs in a way that others don't - and has perhaps rightly decided that this picture needs reforming. However, instead of arguing that one's relationship to one's own memories or beliefs might not best be described in terms of objects to which one has 'access', (a move which may have enabled him to re-characterize beliefs in an enactive way – e.g. as manifest through practice), he has gone the unusual route of leaving this element of the traditional picture untouched and has instead waived the 'privileged' part, so that anyone might have the same 'access' to my own internal 'store' that I have. This would seem to have the unusual repercussion that my beliefs are only 'mine' by virtue of their (literally conceived) spatial location, for in all other respects other people seem to be able to stand in the same relation to my beliefs as I do. I have my beliefs and memories, we might say, in the same sense that I have a wristwatch. Indeed, if we marry this conception with Clark's conjectures elsewhere concerning the plugging in of neural implants to perform mental functions (e.g. Clark & Chalmers, 1998; Clark, 2003a), it seems that others might be able to *steal* my beliefs and memories just as they might steal my wristwatch – or any other piece of my property - by getting hold of my neural implants and plugging them into their own brains.

These two examples, then, evoke a reified conception of internal beliefs and memories as things that one has about one's person ('in the head'), and which another person might gain access to in the same way that I am held to gain access to them. Given that C&C had such a literalistic account of 'internal' mind, we might suggest that the transference of mental properties onto external artefacts was an easy next step for them.

Mental Properties and Mental Property

It will be noticed that in outlining this conception of reified mental properties we have occasionally alluded to the other sort of 'property'. On a Clarkian view, we might say, my relationship to my beliefs, memories etc. is a little like my relationship to personal property. Both are only contingently mine. I might carry such items around with me but someone else could get hold of them. Such a view perhaps brings to mind Hegel's own conception of property which, it will be remembered, also revealed a degree of literalism about vehicles for mental contents. According to Hegel: "I have my life and body as I have other things, only insofar as they express my will" (Hegel, 1821, p.53). In this way his will was also expressed in his property, although it was not "so vividly distinct" (ibid. p.55) as in his body.

The similarity with Clark's account is double-edged. Insofar as Hegel is simply saying that the objects of the world can become extensions of his being he is saying nothing which runs counter to his and Marx's account of externalization and extension elsewhere, which Clark has taken up, in different ways, in his arguments for scaffolding (Clark, 1997) and augmentation of spatial presence (2003a). However, insofar as Hegel has an objectified conception of his being in the world, such that he relates to his body – as a vehicle for his will - in the exact same way that he relates to items of property, then this would seem to replicate C&C's own reified conception of the mental. In both instances, a 'plug-and-play' view of vehicles for mental contents is at work.

Inside Out

We have criticized Clark & Chalmers' 'literally extended mind' then, on the basis that it presents an alienated conception of mind, which conception is co-occurent with a

reified view of non-extended mental contents. We might also criticise it on separate grounds that it is insufficiently dialectical. The thesis of the literally extended mind, as exemplified in the Otto argument, fixates on the idea of mind seeping out into the world, but pays scarce attention to the converse notion of the world seeping into the mind. This oversight seems incompatible with the emphasis given, even in the same paper, on the more fluid, dialectical notion of 'coupling' (C&C, 1998, pp.8-12).

Otto's notebook, it will be recalled, is held to have functional similarity with Inga's memory. In both cases the idea seems to be that Otto and Inga can 'look things up' in their respective information repositories (ibid. p.12). Insofar as this implies a 'storage' conception of memory we might say that objections to such a picture are well known both within the extended mind literature (e.g. Dartnall, as cited in Clark, 2008, pp.97-98) and outside of it (e.g. Toth & Hunt, 1999). However Clark also alludes to another criticism put forward by Chrisley (personal communication), arguing for a disanalogy between biological memory and Otto's note book for developmental reasons:

> …As children we do not begin by experiencing our biological memory as any kind of object or resource…because we do not encounter our own memory perceptually. Instead it is just part of the apparatus through which we relate to and experience the world. Might it be this special developmental role that decides what is to count as part of the agent and what is to count as part of the world? (Clark, 2008, p.101)

Clark is not impressed by this as a line of attack. He suggests that it is possible that a child might begin by viewing her own bodily parts as objects, or conversely that nonbiological cognitive resources might be incorporated into her system at a young age such that they aren't experienced by her as objects; or even that a child might be taught to view her own cognitive faculties in objectified fashion by being "plugged into biofeedback controllers" (ibid). Such developmental possibilities, says Clark, although interesting, are irrelevant to the notebook argument, for "what counts in the end is the resource's current role in guiding reasoning and behaviour, not its historical positioning in a developmental nexus" (ibid. p.102).

It seems possible however that Clark has not dealt with 'the developmental issue' sufficiently here, for – expanding on Chrisley's point – we might say that it is not just that as children we don't experience our memory as an object or resource, but that if we do view or experience it as such later on, this might well be the result of interaction

with external resources. Let us start from the premise that, as Chrisley suggests, the child does not view or experience his memory in a detached, objectified fashion. We might then contrast this with the conception of things according to which we do "look things up in memory" i.e. according to which memory is viewed as a resource to which we stand in a certain relation. It is implied by C&C's own argument that at some point developmentally such a description becomes appropriate. A relatively reasonable developmental hypothesis to derive from this might be that the move from transparent memory to 'looking things up in memory' might not be unrelated to external practices such as "looking thing things up in a notebook" i.e. the development of biological memory, or our view of it, might have its roots in just such practices.

Here we might draw comparison with Vygotsky's notion that "outward, interpsychological relations become the inner, intrapsychological mental functions" (Vygotsky, 1934, p.xxxii). Exactly how this story should be told in relation to objectified memory is open to question. Possibly our very ability to remember at all, beyond a certain degree of competency, is co-dependent on our ability to retrieve information externally. Less radically, perhaps external acts of search and information retrieval scaffold internal acts of search and information retrieval. As Clark says elsewhere, in relation to mental arithmetic: "the trick…is to learn to manipulate a mental model in the same way as we originally manipulated the real world" (Clark, 1997, p.61). At the very least, our conception of ourselves as users of memory as a resource in which things can be 'looked up', seems bound up with our external practices. We only think we can 'look things up' in memory, we might say, because we can look things up in a notebook.

If we accept such an analysis then it seems that C&C's uni-directional notion of an 'extended mind' does not adequately capture the true state of affairs. If anything, the internal cognition of the agent is an extension of notebook bound processes rather than vice versa. The point however is perhaps that, literally speaking, Otto's cognitive processes are no more part of an extended note-book than the notebook is part of Otto's extended mind.

Appendages of the machine

We have so far concentrated mainly on the extended mind hypothesis as it is presented in the Otto argument. In this last part of the chapter we will switch our attention to the 'cyborgian' portion of Clark's work. Although some of the issues this raises have already been broached in our critique of 'reification', and in our suggestion that the latter brings with it a 'plug and play' view of vehicles for mental contents, there are some additional concerns that need to be addressed.

As we noted earlier, in *The Extended Mind* C&C speculate on the use of neural implants for a wide range of tasks (rotating tetris shapes, geometric reasoning, displaying 'memories'). These implants are not typically discussed as examples of mind extension per se, but rather are used to illustrate the argument that if we are willing to consider such devices as part of our cognitive architecture when they are lodged within our nervous system we should also be willing to consider the use of other artifacts as part of cognition when they are external to us. However although not explicitly given as examples of mind extension in the 1998 paper, the implant examples are later utilized by Clark in a further variant of the extended mind argument – perhaps that whose proximity to functionalism is most clear. This is the argument that biological capabilities can be replaced/extended through the incorporation into the agent's body/brain of hi- tech artefacts, with the resulting capabilities being viewed as no less human or cognitive than previously.[15] To some extent this version of the extended mind argument could be viewed as a variant of that which we earlier characterized as the 'figurative' extension of mind, where cognitive capacities are 'extended' though the use of artefacts, via processes such as scaffolding. However there is perhaps a difference here insofar as the cyborgian extended mind (as we shall call it)

---

[15] Dartnall (2004), with reference to an exchange between O'Brien and Clark (O'Brien, 1998) argues for a different distinction here – that between arguments for the extended mind which are dependent on the 'parity principle' and ones which are dependent on the 'complementarity argument': "The parity argument stresses 'the way that extra-neural elements can play a role similar to internal ones'…But [Clark].. provided a 'more interesting and plausible argument' that turns on the way 'external elements may play a role different from, but complementary to, the inner ones'…This is the complementarity argument."

Dartnall thinks that it is the latter style of argument which is developed in (2003). However I am not convinced of the usefulness of this analysis since it turns on an undefined conception of 'sameness' or 'difference' of role (as does functionalism generally.) One man's 'artefact which plays a similar role to internal elements' is another man's 'artefact which plays a different but complementary role to internal elements.'

is less about being an active manipulator of external resources and more about being a passive recipient of technological upgrades.

This variant of the extended mind thesis is developed in most detail in *Natural Born Cyborgs*. Here Clark argues for "the plain literal truth" (Clark, 2003a, p.3) that man is a 'cyborg'. By this he appears to mean not only that human beings have the power to annexe themselves to various external media (artefacts, devices etc) but that realization of human developmental potential is conditional on such mergers. Thus:

> Minds like ours are made for mergers. Tools-R-US and always have been…What makes us distinctively human is our capacity to continually restructure and rebuild our own mental circuitry courtesy of an empowering web of cognition, education, technology and artefacts. (Clark, 2003a, pp.7/10)

In some respects, Clark's position here is not too dissimilar to that of Marx. We have repeatedly stressed that for Marx the self-creation of the human subject is a distinctive feature of labour. Moreover we have also emphasized that how this can involve the use of artifacts which the agent annexes "to his own bodily organs" (Marx, 1867, p.285). Some commentators have taken the latter as evidence of a central concern with "hybrid human-machine embodiment" (Wendling, 2009, p.141), arguing that for Marx "the integration of a set of instruments into the human corporeal schema defines the human species as such" (ibid. p.240). However, unlike Marx, what seems to be lacking in Clark's account of the cyborgian extended mind is any systematic means of distinguishing between human/artefact interfaces that serve to facilitate human development and those that might serve to stifle it.

Reservations

This is not to say that Clark ignores completely the question of possible pitfalls in mind-machine 'mergers'. In the concluding chapter of (Clark, 2003a) a number of worries are raised, all of which are centred on the loosely defined idea of a threat posed by the increasing dominance of technology in human life. Of these the most promising (from the point of view of this thesis) are those that he labels 'Uncontrollability' and, appropriately enough, 'Alienation'. The former is given as the worry that the

dominance of technology, and our integration with it, might result in the loss of control over own lives – that "increased human-machine symbiosis directly implies decreasing control" (ibid. p.175) The latter is given as the worry that "agent technologies may harmfully degrade how people value themselves and treat each other." (ibid. p.178). However, although Clark is prepared to treat such reservations with enough seriousness to spend a chapter on them it is clear that they have no decisive impact on his overall position. Because he has no critique of the ways in which the integration of agent and artefact have been enforced in the past, or the present (e.g. in mechanized labour), he has no principled concern with any future possibilities. It is thus a reassurance for him that:

> ..The kind of control that we, both as individuals and as society, look likely to retain is precisely the kind we always had. (Clark, 2003a. p.175)

Any threat posed by burgeoning cyborgianism can be met provided that we are careful enough and "pay closer attention" (ibid. p.179) to the social implications of new technologies. As a reward we will see the cybernetic equivalent of the end of alienation, the fusion of brain, body and implant, where Marx's 'whole man' (Marx, 1844a, p.91) is replaced by the similarly unitary 'single adaptive unit':

> Our redesigned minds will be distinguished by a better and more sensitive understanding of the self, of control, of the importance of the body, and of the systemic tentacles that bind brain, body, and technology into a single adaptive unit. This potential, I believe, far, far outweighs the attendant threats of desensitization, overload and confusion. (Clark, 2003a, p.179)

Controlling Systems

The problem with this positing of cyborgianism both as a virtue and as a unitary phenomenon, we might say, is that it leaves Clark ill-equipped to analyse factors which determine the direction and flow of control in the agent's various encounters with technology. In particular he is unable to offer any coherent diagnosis of instances where the flow of control is dictated by the artefact itself.

We saw in chapter 2 how Marx described the factory worker's predicament as being that of an "appendage of a machine" (Marx, 1867, p.799). Such a description was given not just for dramatic effect but had a particular analysis behind it. Human beings externalize themselves through productive activity and in so doing realize certain ends, which in a circular way facilitate the production and development of humanity itself. However when productive activity is determined by the machine, teleology also shifts over to the machine's side – the ends realized are the ends of the machine and not the ends of the agent. Rather than the agent realizing herself through productive activity the machine is realizing itself through the productive activity of the agent. Of course, as we said in chapter 2, the machine itself has no intrinsic ends other than those it embodies for a separate class of agents for whom it carries out its operations, and whose knowledge it incorporates (Mclellan, 1977, p.374), so in this sense it is the externalised powers of the controllers/creators of the machine which dictate the activity of the agent.

In contrast with this account Clark in the main treats the technological artefacts with which human agents mesh themselves as neutral enablers of potential. He does not recognize that they are also the products of other human agents and as such incorporate the knowledge and goals of those agents. He thus has little principled sympathy for the argument that, like a machine in a factory, a neural implant is as likely to impose its own logic on the cognitive functioning of the implantee as be a tool for the extension/development of his/her pre-existing cognitive potential. He does not see that this is as much a logical point as a moral worry – that if a neural implant 'extends' anyone's mind, it could just as easily as be that of its designers as that of the recipient.

Transparency

It might perhaps be thought that such difficulties could be circumvented if we pay sufficient heed to the Heideggerian notion of 'transparent equipment' (Heidegger, 1927; see also Clark, 2008, p.10), and indeed, as Mackenzie (2004) also points out, Clark's ideal for enabling artefacts is that they be 'transparent', rather than 'opaque' to the user. Thus he talks of "well-fitted transparent technologies" (Clark, 2003a, p38) and "non penetrative cyborg technologies" (ibid p.28). The suggestion is that the technologies in question should be invisible to the user rather than obtrusive.

Transparent technologies do not disrupt the flow of human practical/cognitive activity but rather "dovetail or interlock with specific cognitive attributes or capacities" (Mackenzie, 2004, p.155). They thus become "integrated so deeply and *fluidly* with our existing biological capacities and characteristics that we feel no boundary between ourselves and the nonbiological elements" (Clark, 2003a, pp.23-4).

This distinction between 'opacity' and 'transparency' perhaps has loose similarities with Marxian conceptions of alienated and unalienated labour. Interfacing with 'opaque' technologies to some degree involves confronting the artifactual world as an 'other', we might say, whereas interfacing with transparent equipment might be characterised in terms of a surmounting of the subject/object divide. The criterion of 'transparency' might therefore, at first glance, seem to offer a possible means of avoiding scenarios where meshing with technology reduces the agent to an "appendage of the machine".

However there is a problem with transparency, which is succinctly articulated by Mackenzie:

> For any particular technology we need to ask: opaque or transparent for whom? ...Transparency and opaqueness are not intrinsic to the technology. (Mackenzie, 2004, p.155)

There are two points here. Firstly a degree of subjectivity is involved in deciding what counts as 'transparent'. One person's idea of a seamless merger might be another's nightmare scenario of technological obstruction or invasion. Here we need look no further than one of C&C's own examples, derived from the 'Terminator' movie, of an inner visual display (C&C, 1998, p.16) which acts as a stand-in for memory. Ignoring for now the philosophical difficulties presented by this arrangement (viewing a series of images does not equate to 'remembering') we might say that whilst such an artefact might be regarded as a suitable, enabling and transparent substitute for biological memory from C&C's perspective, some critics – including this author – might consider it an imposition on the agent's private domain which, if feasible, would disrupt the natural flow of memory and reduce the agent to a passive spectator.

Secondly, this relativity concerning 'transparency' and 'opaqueness' becomes more significant if we consider our propensity to adapt to technology. Under certain circumstances initially obtrusive technologies might come to be experienced as 'transparent' by an agent, in the sense that the agent will get used to them or take them for granted. Such adaptation need in no way be an indicator that the technology is dovetailing in an *enabling* way with the agent's cognitive capacities or attributes. Clark's declarations that "the most seamless of all integrations and the ones with the greatest potential to transform our lives are often precisely the ones that operate deep beneath the level of conscious awareness" (Clark, 2003a, p.34) and that "the trick .. is to acclimatize ourselves to a much more biological relationship with our technologies" (ibid p. 175) are thus far from reassuring, for it is precisely the transformative role of 'unconscious acclimatization' to technological control that is the source of our concern. Marx's labourers, dominated by the mechanized rhythms of mass production, may well come to view this domination 'transparently', insofar as they relate to it unreflectively, but this is not the kind of "fluid integration" with biological capacities that warrants approval or extends cognitive capacities.

Master and Servant

We have said that Clark's conception of cyborgianism does not equip us with sufficient tools to differentiate between meshings with artefacts which extend and enable, and those which stultify development. Neither, we suggested, can the notion of 'transparent equipment' help us here. It will be noted that our critique of Clark's position hinged partly on the idea of a 'flow of control'. We argued that far from being enabling or extensive of the agent's cognitive capacities, mergers with technology can be disempowering when the flow of control is from machine to agent rather than vice versa. As an addendum we might consider the possibility that this issue is not only relevant to the cyborgian version of the 'extended mind'. It will be remembered that Clark & Chalmers considered a further application of the literal extended mind thesis - "socially extended cognition" (C&C, 1998, pp.17-18) - wherein other human beings functioned as extensions of a primary agent. Examples included "the waiter at my favourite restaurant [who] might act as a repository of my beliefs about my favourite meals", as well as "one's secretary, one's accountant or one's collaborator" (ibid) who might play a similar role.

It should be stressed that this suggestion is not central to their account and appears only as part of a stream of speculation towards the end of the paper. Nor, on the basis of it, do we wish to accuse C&C of any heinous anti-Marxian crime. Nevertheless, these remarks certainly demonstrate how questions of mind extension – with or without technological mergers – might overlap with wider issues of power relations. Although C&C, perhaps deliberately, do not confine themselves exclusively to hierarchically oriented examples – allowing in democratic fashion that one's beliefs might also be contained in one's collaborator – their comments could have ramifications if coupled with a belief in the appropriateness of a mental/manual division of labour. The comparison that most suggests itself here, particularly with the examples of the secretary and the waiter, is that of Hegel's (1807) account of the Lord and Bondsman. This fable, as we saw earlier, prefigures Marx's own account of the mental/manual division of labour.

The key issue, for our purposes is that, according to Hegel, in becoming an extension of his master's mind, the bondsman loses his own autonomy. He is reduced to a "dependent consciousness" such that "what the bondsman does is really the action of the lord" (Hegel, 1807, p.114). Here we have the same dynamic that we saw at play in the case of mechanized labour, but without the technological intermediary. The bondsman is an extension of his master rather than the machine, but in both instances the flow of control is from outside in, rather than vice versa. Applied to Clark's case we can thus see how cyborgianism and "socially extended cognition" present us with different manifestations of the same problem. The waiter, the secretary and the neural implantee are not (necessarily) having their minds extended but rather are reduced to extending the minds of others.

Conclusion

In this chapter we have argued that there are different variants of the 'extended mind' hypothesis. We have expressed agreement with certain of these variants, some of which are Marxian in approach and, indirectly, in origin. However we have criticized the

'literally extended mind' as presenting an alienated account of mind, co-occurent with a reified view of 'internal' mind. In addition we have suggested that considerations relating to the origins of our 'notebook-like' conception of memory might pose problems for the unidirectionality of the Otto story. Finally we considered the cyborgian variant of the extended mind hypothesis. This has some parallels with Marxian accounts, but it lacks a principled understanding of the way in which the flow of control from artifact to agent might result in the latter becoming an "appendage of the machine" (Marx, 1867, p.799).

In our next, and final chapter, we will consider a significant precursor to C&C's extended mind hypothesis – Kirsh & Maglio's work on epistemic action.

Chapter 8 - The Epistemic/Pragmatic Dichotomy

Introduction

In the previous chapter we looked at Clark & Chalmers 'extended mind'. In this, our final chapter, the intention is to look at one of the key influences on that account – Kirsh and Maglio's distinction between 'epistemic' and 'pragmatic' action. Kirsh and Maglio's work is of great relevance to this thesis insofar as it provides us with empirical grounds for calling into question the traditional dichotomy between action and cognition – and in so doing recalls Marx's concept(s) of 'praxis'. At the same time however their account is to some degree problematic insofar as it erects a new, and rather confusing, dichotomy in place of the old one – that between epistemic and pragmatic action. Things are further complicated by definitional issues, both within Kirsh and Maglio's account and between their account and Clark's (Clark, 1997; Clark & Chalmers 1998; Clark, 2008) interpretation of it. We will develop these points in the following: In Section 1 we will present a brief summary of Kirsh and Maglio's original (1994) account, along with an equally brief note concerning Clark's (and Chalmers') take on the subject. In section 2 we will consider the relevance of their account to the Marxian analysis given in earlier chapters. In section 3 we will then attempt a loosely 'dialectical' treatment of the concepts of 'epistemic' and 'pragmatic' action, outlining ways in which they might be thought to "interpenetrate" (Engels, 1887, p.35). In Section 4 we will draw some conclusions from this treatment, focussing particularly on problems with 'pragmatic action'. In section 5 we will then look at some additional conceptual issues with the notion of 'epistemic action', before returning to the topic of dialectics in our conclusion.

(I) Epistemic and Pragmatic Action

Kirsh and Maglio

K&M (in 1994) conducted experiments with players of the game Tetris and, from the resultant data, concluded that the strategies adopted by players were not always best explained by the 'planning' perspective of classical AI. According to (at least one version of) a classical model players might be expected to complete a sequence of information processing and computational steps prior to making any move – with the move being thus viewed as the execution of a preconceived plan arising from those steps. However K&M found that players would frequently make moves (rotations or translations of 'zoids') that would be pointless (and unnecessarily costly) if a plan was already in place - moves which did not bring a zoid closer to its final position or orientation. K&M discounted the suggestion that such extra moves might be explained by a revised version of a planning strategy according to which "players begin execution before they have settled on all the details of a plan" (ibid. p.524/5) for, along with other considerations, it was found that in some cases rotations occurred before *any* details of a plan could have been settled on.

K&M hypothesized that the extra moves were not standard 'pragmatic actions' at all i.e. they were not undertaken for the purpose of "bringing a piece closer to its final position" (ibid. p.516). Rather, they were undertaken for a variety of cognitive purposes. As such they were 'epistemic' actions - actions which "are not performed to advance a player to a better state in the external task environment, but rather to advance the player to a better state in his or her internal environment" (ibid. p.541). In the case of zoid rotation the epistemic functions served might include:

> 1 Unearth new information very early in the game
>
> 2 Save mental rotation effort
>
> 3 Facilitate retrieval of zoids from memory
>
> 4 Make it easier to identify a zoid's type
>
> 5 Simplify the process of matching zoid and contour   (ibid. p.527)

In later work (Maglio et al. 2003, 2008) this hypothesis was tested through experiments which established both that presentation of zoids in multiple orientations did lead to faster responses, and that this benefit outweighed the cost (in time) of taking extra actions to rotate the zoids. In short, epistemic actions improved game play.

The existence of such actions, say K&M, undermines the assumption that "the point of actions is always pragmatic" (Kirsh & Maglio, 1994, p.526). In doing so it also undermines the exclusive emphasis on prior planning typical of some versions of traditional AI. In dialectical fashion K&M characterize the latter as "one sided" (ibid) because it "creates an undesirable separation between action and cognition" (ibid) such that "cognition is necessary for intelligent action, but action is never necessary for intelligent cognition" (ibid)

Clark and Chalmers

We have already seen how Clark and Chalmers developed one aspect of this idea. The external rotation of zoids for epistemic purposes was thought to be such a close functional substitute for rotations carried out internally that it seemed reasonable to suggest that the former might also be regarded as part of cognition, and thus was borne the Extended Mind thesis (simplifying the story somewhat!) We will not here consider the soundness or otherwise of such a move, which we have already discussed to some extent in our previous chapter. What we do want to highlight are two examples of Clark's own characterization of epistemic and pragmatic action. The first is to be found in Clark and Chalmers' (1998) paper:

> Epistemic actions alter the world so as to aid and augment cognitive processes such as recognition and search. Merely pragmatic actions, by contrast, alter the world because some physical change is desirable for its own sake (e.g. putting cement into a hole in a dam). (Clark & Chalmers, 1998, p.8)

The second comes from a more recent work by Clark (2008):

> Epistemic actions stand in contrast to pragmatic actions. The latter are actions which are designed to bring one physically closer to a goal. Walking to the fridge to fetch a beer is a pragmatic action. Epistemic actions may or may not yield such physical advance. Instead they are designed to extract or uncover information. Looking inside

the fridge to see what ingredients are available to cook tonight's dinner is a mild species of epistemic action.
(Clark, 2008, p.71)

It is perhaps noteworthy that the above characterizations do not make reference to an 'internal' 'external' divide when distinguishing between the 'epistemic' and the 'pragmatic' – possibly because such a strategy is more problematic if one's starting point is that cognition is spread out across the environment. For purposes of this chapter, however, what is significant is that that the above two characterizations offer some elements not found in Kirsh & Maglio's own definitions. In the former passage there is the suggestion that pragmatic actions are aimed at bringing about physical changes that are 'desirable for their own sake'; and in the latter passage there is the suggestion that any action undertaken for the purpose of extracting or uncovering information, even if this is achieved merely by 'looking', can be considered a species of epistemic action. We will return to these points shortly.

(II) Epistemic Action and 'Praxis'

We have given some indication, then, of the way epistemic and pragmatic action are characterized by Kirsh & Maglio, and also by Clark (and Chalmers). As implied in our introduction, the notion of 'epistemic action', at least as described by Kirsh & Maglio, seems particularly relevant to the concerns of this thesis insofar as it is explicitly presented as a means of confronting a dichotomy (Kirsh & Maglio, 1994, p.526), and the dichotomy in question is that between "action" and "cognition" (ibid). In this way it harmonizes with Marxian concept(s) of "praxis" (Marx & Engels, 1970a), considered both in general terms as the "synthesis of thought and action" (McLellan, 1969, p.10) and more specifically as the notion that ideas are generated from practical activity. Indeed we might say that the concept of 'epistemic action' provides us with a concrete example of how both of these components of praxis can be conjoined. The rotation of Tetris zoids for epistemic purposes facilitates further cognition, but at the same time can itself be viewed as a kind of practical cognition. The latter point, as noted last chapter, has been made by Clark & Chalmers who say of the manipulation of scrabble tiles for epistemic ends (another of Kirsh & Maglio's examples – Kirsh, 1995; Maglio

et al. 1999): "in a very real sense, the re-arrangement of tiles on the tray is not part of action; it is part of thought" (C&C, 1998, p.10).

In thus providing an empirical demonstration of how thought and action might be unified, K&M's account is more radical and thoroughgoing than some Marxian accounts. In particular it will be remembered that the unification of theory and practice which Jha attributed to the Italian Marxist Labriola (1897) involved only a "cyclical movement of theory and practice" (Jha, 2010, p.208). This suggests a less than dialectical unity, reminiscent of Merleau-Ponty's "mechanical action" (Merleau-Ponty, 1942, p.160) or Clark's "catch and toss explanation" (Clark, 1997, p.105) albeit in relation to a different domain of explanation. In contrast K&M's account promises something beyond mere "interleaving" (K&M, 1994, p.524), a genuine synthesis of action and cognition in the form of epistemic action.

Conversely, insofar as K&M do conceive of thought and action in separation, it is in a way which chimes well with Marxian critiques of idealism. The objection to the latter was that it emphasized "thought before the fact" (Labriola, 1897, p.10), or explained "practice from the idea" (Marx & Engels, 1970b, p.38). This critique is mirrored in K&M's own critique of classical AI (as cited above) wherein "cognition is necessary for intelligent action, but action is never necessary for intelligent cognition" (K&M, 1994, p.526). We note too that the rejection of the necessary priority of cognition brings with it a concomitant rejection of the necessity of prior planning to intelligent action. The account is thus in accord with our own interpretation of the spontaneous Marxian agent who must, in certain circumstances, be able to 'act before he thinks'.

(III) The non-exclusivity of the epistemic and the pragmatic

However, having argued that the concept of 'epistemic action' provides us both with a praxis-like unity and a critique of idealism we have also to recognise that Kirsh & Maglio's (and Clark's) analysis presents us with some difficulties. One significant issue lies in the fact that, having posited this unity, Kirsh and Maglio then proceed to characterize it as one pole of a dichotomy, the occupant of the other pole being

'pragmatic action'. Perhaps predictably from a dialectical perspective, it would appear that the two poles of the dichotomy "interpenetrate" (Engels, 1887, p.35) in different ways. Although Clark (2008) alludes to this when he says (above) that epistemic actions "may or may not" bring one closer to a goal, and the issue is also briefly touched upon by Kirsh (2006), it has not yet been discussed at any length in the literature, and neither, we would contend, has the extent of this interpenetration been recognised. In the following we will attempt to redress this balance in Hegelian fashion by outlining different ways in which 'epistemic action' and 'pragmatic action' seem to overlap or intermingle. En route we will also highlight some other possible difficulties with the notions of 'epistemic' and 'pragmatic' action, as they arise.

1) Epistemic actions are also performed to bring one closer to goal state.

In (K&M, 1994) pragmatic actions are defined as "actions whose primary function is to bring the agent closer to his or her physical goal", a definition that is implicitly echoed most recently in Maglio et al.'s (2008) account, where it is said that *epistemic* actions "are used to simplify internal computations *rather than to move close to an external goal state.*" (my italics). So, pragmatic actions have something to do with getting oneself closer to one's physical goal state. At the risk of sounding pedantic, however, it should perhaps be pointed out that in a task bound environment any action motivated by the desire to complete the task can be characterized as having the purpose of bringing the agent closer to his or her physical goal. This would include those actions characterized as 'epistemic'. Why, for instance, does the Tetris player carry out zoid translations which move the zoid physically further away from the goal? To help him/her verify the column placement of the zoid (K&M, 1994, p.539). But why is it important to get one's bearings in this way? So that one can then place the zoid in the appropriate place (i.e. move to goal state.) Thus, on at least one interpretation of the phrase, the 'primary function' of an epistemic action is also to bring the agent closer to goal state. Indeed, as Clark (1997) also points out, K&M not only acknowledge this but go further by suggesting that 'state space' might be redefined to include both physical states and informational states (K&M, 1994, p.515). On this analysis, a

relevant change in informational state arising from performance of an epistemic action presumably *does* bring one closer to goal state as measured by progress along the nodes of a revamped state-space graph.

Bearing this in mind we could perhaps modify our definition of pragmatic action accordingly – pragmatic actions are those whose primary function is to bring the agent *directly* closer (in a non-informational sense) to his/her goal whereas epistemic actions are those taken for epistemic purposes in order *eventually* to reach that same goal.

As an aside, however, we perhaps also ought to note a further complication concerning what counts as 'moving directly closer to goal state', one which has nothing to do with epistemic or informational considerations. In the Tetris example, and in Clark's 'walking to the fridge to fetch a beer' example, moving directly closer to goal state correlates well with moving physically closer to a particular configuration or locale, but this need not always be the case. In certain problem-solving scenarios – such as the Tower of Hanoi problem – one might need to move physically further away from a state of affairs that resembles goal state (for non epistemic reasons) in order to get closer to goal state. Likewise we might imagine that Clark's fridge was placed in a maze so that one had to move further away from it in order eventually to reach it. Thus 'moving closer to goal state' may sometimes be better understood as 'making progress in a sequence of physical steps necessary to bring one to goal state'. (We will not explore this further here.)

2) Epistemic actions have their own goals?

The argument given above (in 1) is that epistemic actions are undertaken to bring one (indirectly) closer to the self-same goal that pragmatic actions are undertaken to bring one ('directly') closer to. So far, so uncontroversial. However it also seems that epistemic actions might bring one directly closer to a goal – albeit a sub-goal in relation to the ultimate goal of the pragmatic action. When I look inside the fridge to see what ingredients are available before I make a meal, my immediate goal is to see what ingredients are available and this is achieved by looking in the fridge. This kind of overlap seems to be acknowledged by Kirsh in a later paper where he says that

"epistemic actions could be seen as pragmatic with respect to advancing epistemic ends when these were the external goal" (Kirsh, 2006, p. 252). If an action which is epistemic in relation to a particular pre-defined goal can count as pragmatic in relation to its own separate sub-goal, could we then say that there is a degree of relativity concerning what counts as a pragmatic or epistemic action?

In attempting to answer this question we come across what appears to be a significant problem, for it is not clear what characteristics of epistemic and pragmatic action are essential for them to count as such. In K&M's original (1994) paper the distinction was tightly linked to a particular micro-story. Epistemic actions in Tetris were seen as possible replacements for internal operations. On a "classical information processing account of Tetris cognition" (ibid. p.518) a certain sequence of operations (early vision, attention-directed encoding, generate and test, motor planning and control) was thought to be a necessary pre-requisite for action. K&M's revised account replaces or supplements elements of this sequence with external operations (e.g. zoid rotation) such that there is a "tight coupling" (ibid. p.542) between internal cognition and external manipulation, with a resultant increase in speed of processing.

The 'micro-epistemic actions' of Tetris, then, are put forward as part of an explanation of the underlying structure of intelligent action. It would seem misleading, we might suggest, to characterize these actions as literal realizers of their own epistemic goals, for by their nature they cannot be underpinned by the modelling of a goal state. The agent does not rotate the zoid (when doing so for epistemic purposes) with a predefined configuration in mind for if s/he did so this would defeat the object of the action in the first place – which is to bypass the need for extensive internal planning by making use of external resources, thereby speeding up processing. Whilst it might be possible to talk in a figurative way of such actions as achieving 'epistemic goals', their *modus operandi* is actually the opposite of this, for they are precisely not underpinned by planning, goal-modelling cognition. They are not underpinned by anything for they are *replacements for* (steps in) an internal cognitive sequence that was itself previously thought to underpin action.

However, we note that although epistemic action is *linked* to this kind of story in Kirsh and Maglio's work there is nothing in either their or Clark's definitions which makes

such an interpretation necessary. Whether they are actions "an agent performs to change his or her computational state" (K&M, 1994, p.514), actions which "alter the world so as to aid and augment cognitive processes such as recognition and search"(C&C, 1998, p.8), or actions "designed to extract or uncover information" (Clark, 2008, p.71), there is no explicit ruling regarding *how* such actions are to be carried out. (Nor for that matter is there any suggestion that such actions need to be ones which *replace* operations that could plausibly be carried out 'internally' on a classical account - a point we shall return to.) Hence the door seems to be open to include actions which are underpinned by traditional planning style cognition. Looking in the fridge to see what ingredients are available to make a meal meets the loose definitional criteria for what counts as an epistemic action (or at least Clark asserts that it does) but might be satisfactorily implemented by a classical goal modelling agent. Likewise, the assertion that memory saving strategies such as placing a key in a shoe (Kirsh & Maglio, 1994, p.515; Kirsh, 2006, p. 252) count as epistemic actions, implies nothing about a requisite underlying structure for such actions. Presumably they could therefore be realizers of epistemic sub-goals in a more than figurative sense, unlike their micro-counterparts.

3) (Almost) any action might have epistemic pay-off

In the 1994 paper Kirsh and Maglio make the following observation:

> From a methodological standpoint, it is often hard to prove that an agent performs a particular action for epistemic rather than for pragmatic reasons, because an action can serve both epistemic and pragmatic purposes simultaneously. Rotating a zoid in the direction needed for final placement may also help the player identify the zoid. This frequently makes it difficult to quantify the relative influence of epistemic and pragmatic functions (Kirsh & Maglio, 1994, p.527)

In fact K&M may be underplaying the extent of the difficulty here for it seems arguable that most actions in Tetris provide some kind of epistemic benefit. Each rotation or translation of a zoid whether or not it brings us physically closer to our goal can be said to have epistemic pay off insofar as it provides the agent with easy visual access to an updated physical state space. The resultant configuration will help the player decide what to do in the next move. Moreover it will do so not just by supplying abstract

information about the state space (which could be had non perceptually) but by providing a perceptual guide for the next move.

It seems unlikely that Kirsh and Maglio would disagree with this. Would they then want to claim that most actions are epistemic actions? Perhaps they might want to avoid this by making a distinction between actions which have epistemic pay-off and 'truly' epistemic actions by reference to the purpose for which the action was carried out. Such a distinction seems implicit in the above passage (although "serving a purpose" is perhaps more passive than doing something for a purpose) and is expressed more clearly in a remark immediately following that passage – "there are clear cases where the only plausible *rationale for a particular choice of action* is epistemic."(ibid.) It also seems implicit in purpose-laden language elsewhere in the paper:

> Epistemic actions – physical actions that make mental computation easier, faster or more reliable – are external actions that *an agent performs to* change his or her computational state. (Kirsh & Maglio, 1994, p.514, my emphasis)

and we should note, too, definitions provided in (Maglio et al. 2003, 2008) which both describe epistemic actions as "physical actions *people take to* simplify internal problem solving." (p.1 in both cases, my emphasis)

There is the suggestion, then, that epistemic actions are motivated by epistemic *purposes* and we could conclude from this that not all actions which have an epistemic payoff are thereby epistemic actions. However there is perhaps a difficulty here which K&M themselves hinted at when they stated that 'it is often hard to prove that an agent performs a particular action for epistemic rather than for pragmatic reasons'. As Destefano et al. put it, in a separate study of strategies adopted by Tetris players, "we have no access to an individual's intentions" (Destefano et al., 2011, p.2709). Suppose we agree that a particular action in Tetris has epistemic payoff but we want to ask the further question of whether the epistemic payoff was deliberately aimed at before we are happy to baptise it as a truly 'epistemic action'. How are we to go about ascertaining this? Are we to ask the players themselves? This does not seem to be a reliable strategy for there is no reason to suppose that players need be able to provide reasons for a particular move/set of moves retrospectively. We note in particular, that speed of play means that a certain degree of automaticity needs to be cultivated.

Moreover at the micro-level, as we have seen, part of the reason for singling out 'epistemic' actions as distinctive is that, in some cases at least, they are undertaken before any internal deliberation could plausibly have been carried out (and indeed – on an extended mind analysis – they are thought to *constitute* part of the deliberation.)

We have, of course, the suggestion that more experienced players tend to make more 'epistemic' (i.e. epistemic only) moves (Maglio & Kirsh, 1996 - although this is to some degree disputed by Destefano et al., 2011). Insofar as this is the case it could be taken as implying the existence of a purposive strategy to improve game play. But again we might point out that if there is such a strategy it need not be one that could be consciously articulated by the agent. In fact, to an extent the proof of the pudding is taken to be in the eating. If an action is undertaken deliberately and has epistemic pay off which is relevant to and improves game play, it is assumed to be an action undertaken for epistemic purposes. We have, what might be called, 'pragmatic' (in a traditional philosophical sense) criteria for the individuation of epistemic actions. There is nothing wrong with this but it is not clear that such criteria can supply us with a means of differentiating between 'truly' epistemic actions and actions which 'merely' have epistemic pay-off.

4) Are there any purely pragmatic actions?

If we accept that actions which take us closer to goal state, but which also have epistemic pay off, might constitute a sub-species of epistemic action then it would seem to follow that (in the game of Tetris at least) there may be no purely pragmatic actions. We said that "most actions" in Tetris are likely to have some epistemic payoff. In fact the only exception to this would seem to be the final move (in a game or a sequence of moves) – the dropping of a zoid into a slot. Here it perhaps could not be claimed that some goal-relevant epistemic advantage is obtained because the agent has already arrived at goal state (although this is not to say that there might not be epistemic pay off in relation to wider goals e.g. improving game play.) It seems problematic, however, that this is the only example of a purely pragmatic action which we can muster, for then 'purely pragmatic action' would seem to have become a synonym for 'arrival at goal state'.

5) Levels of action

A further consideration – perhaps implicit in some of what has already been said – is that of 'levels of action'. Without wishing to endorse any particular theory concerning the individuation of actions (see e.g. Goldman, 1970; Danto, 1973; Honderich 1988) we might say that there is an uncontroversial sense in which some actions can be said to be made up of other actions. For instance an action described as "switching the light on" might contain the sub actions of "raising an arm" and "flicking the switch". Bearing this in mind we might note that the category used to describe an action at a macro level ('epistemic' or 'pragmatic') may not be applicable to all the actions that compose it. An action described as 'pragmatic' at the macro level might contain many 'epistemic' actions at the micro level. Likewise, an action described as 'epistemic' at the macro level, might contain many 'pragmatic' actions at the micro level.

As an instance of the former let us take Clark and Chalmers' example of a pragmatic action given in the Extended Mind paper (1998, p.8) – "putting cement into a hole in a dam." This is contrasted with epistemic rotations of Tetris zoids or movement of scrabble pieces as an action which "alters the world because some physical change is desirable for its own sake" (ibid). However we might suggest that "putting cement into a hole in a dam" could also be considered as a macro action, perhaps on a level with playing a whole game of Tetris or Scrabble, and as such could be viewed as made up of smaller sequences of actions. From this point of view there is no reason to think that some of the actions which make up the cement story should be any less epistemic than the various shape rotations in Tetris that K&M draw attention to. The hole filler – in manipulating the cement – may try various strategies which, although they do not bring him/her physically closer to the goal state, will help him/her to get a better grasp of the problem space e.g. toying with cement already in the hole may help to guide the agent in deciding where to place subsequent pieces of cement.

As regards epistemic actions containing non-epistemic actions – here we could imagine a poorly designed mechanical version of the game of Tetris where three lever moves are required to rotate the shape through 90 degrees - the first lever move sets the direction of the rotation, the second sets the degree of the rotation and the third initiates

the rotation itself. The agent may then rotate the zoid for epistemic purposes – but his/her action will be composed of three sub actions which are not undertaken for epistemic purposes.

(IV) Incommensurate Categories

We have argued that there are various ways in which 'epistemic action' and 'pragmatic action' seem to intermingle or overlap – (1) epistemic actions are also undertaken to bring one closer to the self-same goal that pragmatic actions are undertaken to bring one closer to; (2) epistemic actions (loosely defined) bring one closer to their own goals; (3) most actions (at least in Tetris) have epistemic pay-off (thereby (4) making it hard to find a 'purely pragmatic' action); (5) epistemic actions contain pragmatic actions and pragmatic actions contain epistemic actions.

How, then, are we to analyse these results? Kirsh himself recognizes that the distinction between epistemic and pragmatic action "is not hard and fast" (Kirsh, 2006, p.252) but the degree of intermingling and overlap we have described seems to suggest more than this. On the other hand, it does not seem appropriate to attempt a straightforward dialectical resolution here. For one thing, as we suggested earlier, if there is a dialectical 'unity' to be found in Kirsh and Maglio's work it is that which is already contained within the concept of 'epistemic action' i.e. the synthesis of action and cognition.

One thought that might strike us, particularly as regards (1) and (2), is that of whether the categories of 'epistemic' and 'pragmatic' action are really the right sort of things to constitute a dichotomy at all. Perhaps, despite being set up as antinomous, they could never be "two poles of an antithesis…inseparable as they are opposed" (Engels, 1887, p.35) because they are incommensurate categories. It seems possible that *some* of the overlap between epistemic and pragmatic action arises because 'pragmatic' and 'epistemic' are not both 'types of action' in the same sense, but are rather more like answers to two different types of question about action. One might imagine that a

theorist had come up with the following classification for actions – 'actions that one undertakes whilst wearing shoes' and 'actions that one undertakes when on a train'. When one points out that many actions that one undertakes on a train are also actions that one undertakes when wearing shoes the theorist will admit that there is some overlap in the real world but will say that nevertheless the distinction can still be made in principle. One would be left with the nagging feeling that there is something fundamentally wrong with this reply. It is not merely that there is some overlap. 'Actions undertaken on a train' is an answer to the question 'Where were the actions undertaken?' whilst 'Actions undertaken whilst wearing shoes' is an answer to the question 'What were you wearing on your feet when you undertook the action?' There is no useful material for a dichotomy here – even if classical theorists had traditionally conceived of all action as taking place on a train.

Can something similar be said of the distinction between epistemic and pragmatic actions? As we saw, according to C&C:

> Epistemic actions alter the world so as to aid and augment cognitive processes such as recognition and search. Merely pragmatic actions, by contrast, alter the world because some physical change is desirable for its own sake (e.g. putting cement into a hole in dam). (Clark & Chalmers, 1998, p.8)

This reference to a physical change that is 'desirable for its own sake' has a strange ring to it. Firstly, as alluded to earlier, it is something of a departure from the account given by Kirsh and Maglio which does not make any reference to the desirability of the consequences of a pragmatic action – just that it is carried out to bring one physically closer to a goal. Secondly it is surely often a matter of interpretation in any particular case whether an action is 'desirable for its own sake' – putting cement into a hole in a dam, for example, would not seem to be something that I would wish to do for its own sake, but rather something I would do to stop the dam leaking. Perhaps eating a chocolate cookie or having sex are actions that I might take because they are desirable for their own sake, but then pragmatic actions would suddenly seem to have become linked with hedonistic actions and it seems unlikely that this was Clark and Chalmers' intention.

We do not wish to be overly flippant here of course, and nor do we wish to read too much into one particular turn of phrase which Clark and Chalmers might well admit could be improved. On the other hand this attempt to express informally what pragmatic actions are about does perhaps bring to the fore a certain disparity between the two classes of action, for what motivates Clark and Chalmers' suggestion here might be the thought that pragmatic actions are just actions 'per se'. It is 'epistemic action', we might say, which wears the trousers in the epistemic/pragmatic dichotomy. An interesting class of actions called 'epistemic actions' has been unearthed and these need to be contrasted with non-epistemic actions of some sort. Hence epistemic actions are described as actions that are carried out to get some epistemic pay off whereas pragmatic actions are just carried out. However we might argue that most actions are likely to have been carried out for a purpose, epistemic or otherwise, and so this does not seem to be a true dichotomy. Rather what the distinction boils down to is, on the one hand, actions whose purpose is epistemic, and on the other hand, action when you aren't talking about purposes at all. In this sense, epistemic action provides one answer to the question "why did you perform that action?" whereas the category of 'pragmatic action' is used merely to specify what the action is (an answer, perhaps, to the question "what action did you perform?").

Things seem a little more complicated if we turn from Clark & Chalmers' to K&M's definition of pragmatic and epistemic action, for here both types of action seem, on the face of it, to be offering explanations in terms of purpose and hence to be answering the same question i.e. 'Why did you perform that action?'. In the case of pragmatic action the answer is 'to get me physically closer to a goal, and in the case of epistemic action it is 'to achieve some epistemic pay off (which ultimately will help me get closer to my goal)'.

However perhaps there is some equivocation with the notion of 'purpose' here. Epistemic actions, as we suggested earlier can also be said to be undertaken for the purpose of bringing the agent physically closer to a goal, either directly or indirectly (depending on what sort of epistemic action we are talking about.) In fact, *insofar as we adhere to a goal oriented conception of action* then the 'purpose' of all action is to bring us closer to a goal. Thus this seems a very impoverished sense of 'purpose', one which provides a restricted – almost tautological – answer to the question 'why did you

perform that action?', and one which cannot be used to individuate a particular class of action per se. We might contrast it with the richer sense of 'purpose' implicit in asking the further question 'Yes, but why did you want to get closer to that goal?' We note that the answer to this might be 'for epistemic reasons', but it might equally as well be 'because someone told me to', 'because I like sandcastles' or a host of other answers:

	Question	Answer	Applies to:
1	'Why are you performing that action'?	To get closer to a goal.	All actions including epistemic.
2	'Why do you want to reach the goal?'	'For epistemic reasons' 'Because someone told me to' etc	All types of action e.g. epistemic

The first question (on the K&M definition), then, does not provide us with an answer which specifies a particular class of actions. All actions (including epistemic ones) are performed to bring one closer to a goal, if one has a goal oriented conception of action. The second question provides us with a host of possible answers and thus a host of possible types of action, one of which is epistemic. Hence on this analysis we might be tempted to drop the notion of pragmatic action completely as it does not seem to designate a separate class of actions. Rather, as before, it just seems to denote 'action' per se, but considered from the perspective of goal fulfilment.

Nevertheless – we must be wary of throwing the baby out with the bath water. There is nothing wrong in deciding to apply the term 'pragmatic' to particular actions in certain circumstances where rules for its use are clear. In the toy scenario of Tetris - where the goal is clearly specifiable, where all action is ultimately motivated by the desire to reach that goal, and where each discrete action results in a physical state of affairs that is either closer to or further from that goal – K&M have shown that it is useful to be able to distinguish between actions which take us physically closer to the goal and

those that don't, and then to investigate the reasons for actions of the latter sort. In this scenario the former type of action might as well be labelled 'pragmatic' as anything else. As we implied earlier (in section III 2)) such actions can be seen as 'pragmatic' relative to a certain predefined goal which epistemic actions are not (directly) 'pragmatic' relative to. Moreover, on one interpretation of 'epistemic action', there is also the suggestion that the micro-epistemic actions of Tetris might not be realizers of their own sub-goals, and so are not *directly* 'pragmatic' relative to anything. What this does not seem to give us however is a meaningful non-relative notion of 'pragmatic action' which could have general applicability in the real world, such that it does not just mean 'action'.

(V) The Generality of the Epistemic

In the above section, then, we have suggested that insofar as 'pragmatic action' is identified with getting closer to a particular goal state, or bringing about changes which are desirable for their own sake, it is difficult to see how it can name a distinct type of action, outside of certain well specified toy scenarios. However we also face some problems with epistemic action. We saw (in section III) that in Tetris it was not easy to find examples of non epistemic actions if we allowed that any action with epistemic pay-off might count as an epistemic action. Once such a concession is made the problem would seem to be that, if almost any action might count as epistemic, then the category of 'epistemic action' loses much of its usefulness. It is open to question whether the considerations relating to the difficulty of ascertaining the purposeful epistemic intent of actions in Tetris are convincing, and if so, whether they also apply to actions in the world outside of Tetris. We can perhaps argue that it is less difficult to ascertain the intentions of an individual who places a key in her shoe, than one who rotates a zoid at speed. There is thus a sense in which everyday epistemic actions present less of a problem in this respect – although we might also acknowledge that there is a trade-off here, since there is no requirement that the latter be underpinned by a distinctive non goal-driven architecture.

However, there are other considerations which might make it difficult to place limitations on the number of actions which qualify as 'epistemic'.

266

Substitutability

As remarked earlier (III – 2), epistemic action as it is presented in K&M's account is linked with certain features which do not form part of its definition. One such feature is 'substitutability'. This is the idea that the action undertaken replaces an operation which could have been carried out internally e.g. internal rotations of zoids are replaced by external rotations of zoids. Whilst substitutability plays a significant role in K&M's account it is not clear whether this is felt to be a necessary component of the 'epistemic-ness' of an action or whether the direct substitutability of external for internal operations merely serves as a convenient illustration of the difference between two approaches to intelligent action. We might note that recently Kirsh seems to have abandoned the idea of substitutability, arguing that there are "epistemic activities we can perform outside that we cannot duplicate inside" (Kirsh, 2010, p.452). This is perhaps in keeping with that version of the extended mind thesis which posits a figurative 'extension' of cognitive capacities through the manipulation of external resources, rather than a literal extension of mind in space. However it is perhaps significant that in the same paper Kirsh seems also to have abandoned talk of 'epistemic actions' as a narrowly defined class of action – preferring to talk more loosely of "epistemic activity" (ibid. p.449) and "epistemic interaction" (ibid. p.454).

What *is* clear is that substitutability does not seem to be important to Clark's (2008) account, an account which is very much concerned with 'epistemic actions' as a distinct class of actions. Whilst the idea of substitutability as outlined in K&M's original (1994) paper seems to have been a key impetus in the development of Clark & Chalmer's (1998) extended mind thesis, all that is required for an action to be epistemic on Clark's later account is that it is "designed to extract or uncover information" (Clark, 2008, p.71). Hence, as we have seen, "looking inside the fridge to see what ingredients are available" (ibid) serves as an example (albeit a 'mild' one) of such an action. There is no suggestion that 'looking inside the fridge' need be a substitute for an operation that could be undertaken without looking in the fridge. It is not implied, for example, that on a classical model the agent would be expected to remember what was in the fridge without looking inside it.

The problem here would seem to be that once epistemic action is generalized to include any kind of action undertaken to extract information it has become a much vaguer kind of notion – from this perspective, for example, science could be seen as a huge accumulation of epistemic actions.

Perception

Moreover if all that it takes for an action to be epistemic is that it uncovers information, there is the additional question of what we are to say about perception. Most action (traditionally conceived) is accompanied by perception, and most perception is likely to involve extracting relevant information.

It might perhaps be objected that simple acts of perception are ruled out definitionally from being epistemic actions because the latter must have an effect on the local environment - they need to "change the world" (K&M, 1994, p.513) or "alter the world" (C&C, 1998, p.8) in some way. However a problem we face here is that of establishing what counts as changing the world. A fan of embedded/embodied cognition might argue - with some justification - that the agent is him/herself a part of the world. In moving his head to inspect the surrounding environment he is therefore altering the environment. Indeed, whilst K&M state in their original paper that their interest is with "ordinary actions not sensor actions" (K&M, 1994, p. 515) Clark's own example of looking in the fridge seems to imply that mere 'looking' is substantial enough an action in itself to count as epistemic (although there is the possibility that he intended 'opening the fridge door' to be included as part of the action.)

If 'looking' counts as an epistemic act are we then to say that most action is epistemic? As before this would seem to undermine the usefulness of 'epistemic action' as a category. Moreover it also makes it harder to differentiate between explanations in terms of epistemic actions and explanations which take a more classical approach according to which plans are updated via sensory input. What is to stop us from describing the 'sense-model-plan-act' cycle (Brooks, 1991) as 'looping into the world' at the sensing stage with sensing itself now recast as an 'epistemic action'?

Our problem, then, would seem to be that loose definitions tend to make the notion of 'epistemic action' too general to be helpful. If we allow that an epistemic action can be any action which is designed to uncover information – whether or not that information could have been uncovered by non-active (i.e. internal) means – then the gate is open to include simply 'seeing something' as an example of epistemic action.

Conclusion - Back to Dialectics?

The foregoing analysis perhaps places us in something of a quandary. Our earlier characterization of epistemic action (in Section II) was as a praxis-like 'unity' of action and cognition. This characterization proceeded from a wider dialectical perspective according to which one would expect action and cognition to be everywhere inter-related rather than rigidly demarcated. Such an outlook also motivated our analysis in section III, 3, where it was suggested that it was difficult to find examples of non-epistemic actions in Tetris if we allowed that any action with epistemic pay-off might be a subspecies of epistemic action. Likewise it was such a perspective which informed out argument in section III, 5. There we said that epistemic macro actions could contain 'pragmatic' micro actions and vice-versa. Although we later argued that the notion of 'pragmatic' action was to some degree problematic anyway, we can still make sense of that argument as asserting a generalised interdependence between the epistemic and the non-epistemic.

However, in section 5, we have also complained about the loose way in which the term 'epistemic action' is sometimes utilized. The underlying suggestion here is that more precision is needed if we are to retain what is distinctive in the phenomenon described in Kirsh & Maglio's original paper. Thus it would appear that we are being pulled in two opposing directions. The dialectical portion of our analysis wants to generalise the epistemic dimension of action, whereas the analytical portion of our work wants clear demarcation between 'epistemic action' and other types of action, so that 'epistemic action' remains a useful category. Is it possible to reconcile these two tendencies? Perhaps an attempt might be made as follows:

It is true, we might say, that action and cognition are intertwined with each other as the general case. It is not just that "epistemic actions are everywhere" (Kirsh, 2006, p.252) but that cognition and action are coupled with each other, and the world, in an ongoing fashion. This does not mean, however, that we cannot isolate instances where external resources are utilized in a distinct or striking way, but it is perhaps misleading to present such instances as if they were hermetically sealed off from an otherwise non-epistemic involvement with the world.

Conclusion

In our exploration of the 'retreat from alienation' in cognitive science we have considered a diverse range of themes and issues. It is hoped, nevertheless, that the reader will see some underlying unity in this disparate collection of ideas. Indeed it is the notion of 'unity' itself which informs the majority of our discussions – a unified conception of the agent, facilitated by anti-dichotomous reasoning, as part of a broader dialectical, and hence unificatory, outlook.

There is much that we have missed out of our analysis. One significant omission is perhaps a detailed comparison of our own Marxian take on cognitive science with contemporary Heideggerian accounts of cognitive science (e.g. Wheeler, 2005). Such a comparison would entail a summary of relevant parts of Heidegger's work, considered in relation to Marx's work, with some acknowledgement of areas of overlap (including the possibility of a 'Heideggerian Marxism'- Wollin & Abromeit, 2005); and the application of the resultant insights to the arena of cognitive science. Whilst such an exercise would unquestionably be useful, and might resolve some of the issues left hanging in our own analysis, space considerations made it impractical.

Nevertheless, even with such omissions, it is apparent that we have only managed to scratch the surface with the majority of the issues raised. The thesis should therefore be viewed only as an initial foray into territory that needs to be explored in a great deal more depth.

_____

Bibliography

Acton, H.B. (1967). 'Hegel' in Inwood, M.J. (ed), *Hegel – Selections*. London: Collier Macmillan. pp. 1-34.

Adams, F. & Aizawa, K. (2010). *The Bounds of Cognition*. Oxford: Wiley-Blackwell.

Alac, M., & Hutchins, E. (2004). 'I See What You are Saying: Action as Cognition in fMRI Brain Mapping Practice.' *Journal of Cognition and Culture*, 4(3): pp. 629-661.

Aleksander, I. (2005). *The World in My Mind, My Mind in the World*. Exeter: Imprint Academic.

Anderson, M.L. (2003). 'Embodied Cognition: A Field Guide', *Artificial* Intelligence 149 (1), pp. 91 – 130

Arthur, C.J. (ed. 1970), *The German Ideology*, London: Lawrence & Wishart: revised edition 1974. Editor's Introduction, pp. 4-34.

Arthur, C.J. (1983), 'Hegel's Master-Slave Dialectic and a Myth of Marxology' in *New Left Review*, Nov-Dec 1983, pp. 67-75.

Arthur, C. J. (1986). *Dialectics of Labour: Marx and his Relation to Hegel*, Oxford: Basil Blackwell.

Arthur, C.J. (2004). *The New Dialectic and Marx's Capital*. Leiden: Brill.

Avineri, S. (1970). *The Social and Political Thought of Karl Marx*, Cambridge: Cambridge University Press.

Avineri, S. (1972). *Hegel's Theory of the Modern State*, Cambridge: Cambridge University Press.

Ayer, A.J. (1954). 'Can There Be a Private Language?', *Proceedings of the Aristotelian Society, Supp. Vol 28*. Reprinted in *Pitcher, G.* (ed. *1966*). *Wittgenstein, The Philosophical Investigations*. London: Macmillan. pp.251-266.

Ayer, A.J. (1956). *The Problem of Knowledge*. Harmondsworth: Penguin Books.

Bakunin, M. (1882). *God and the State*. 1970 edition. New York: Dover.

Barbrook, R. (2007). *Imaginary Futures*. London: Pluto Press.

Barnes J. (1987). *Early Greek Philosophy*. Harmondsworth: Penguin.

Baron-Cohen, S, Tager-Flusberg, H, & Cohen, D, (eds. 2000). *Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience.* Oxford: Oxford University Press.

Baumgartner, P., & Payr, S. (eds. 1995*). Speaking Minds: Interviews with Twenty Eminent Cognitive Scientists*. Princeton: Princeton University Press.

Blackmore, S.J., Brelstaff, G., Nelson, K. & Troscianko, T. (1995) 'Is the richness of our visual world an illusion? Transsaccadic memory for complex scenes'. *Perception*, 24, pp. 1075-1081

Bernstein, R. (1999) *Praxis and action: Contemporary Philosophies of Human Activity*. Philadelphia: Pennsylvania Press.

Boden, M.A. (2006). *Mind as Machine*, Oxford: Clarendon Press.

Bottomore, T. (ed. 1991). *A Dictionary of Marxist Thought*. Oxford: Blackwell.

Bopry, J., *et al*. (2002). 'Francisco J. Varela 1946-2001'. Special Issue of *Cybernetics & Human Knowing,* 9 (2).

Braverman, H. (1974). *Labour and Monopoly Capital*, Monthly Review Press, New York

Brooks, R. (1991a). 'Intelligence without Reason', Proceedings of the Twelfth International Joint Conference on Artificial Intelligence, pp. 569-595.

Brooks, R. (1991b). 'Intelligence Without Representation', in *Mind Design II*, Haugeland, J. (ed.), Cambridge, Massachusetts: MIT Press.

Burbidge, J.W. (2006). *The Logic of Hegel's Logic*, Canada: Broadview Press.

Campbell D.T. (1974): ''Downward causation' in Hierarchically Organized Biological Systems', in: *Studies in the Philosophy of Biology*, F.J. Ayala & T. Dobzhansky (ed.), (Macmillan Press), pp. 179-186

Chrisley, R. (1995) 'Weak Strong AI: An Elaboration of the English Reply to the Chinese Room,' Pylkkänen, P. and Pylkkö, P. (eds.) *New Directions in Cognitive Science: Proceedings of the International Symposium*, Saariselka, 4-9 August 1995, Lapland, Finland. Helsinki: Finnish AI Society, pp 178-182

Chrisley, R. (2003). 'Embodied Artificial Intelligence', in *Artificial Intelligence*, vol. 149 pp. 131-150.

Clark, A. (1997), Being There: Putting Brain, Body, and World Together Again, Cambridge, Massachusetts: MIT Press.

Clark, A. & Chalmers, D. (1998), 'The Extended Mind', *Analysis* 58(1), pp. 7-19.

Clark, A. & Chalmers, D. (1998a), 'The Extended Mind', slightly different version, accessed online at http://consc.net/papers/extended.html, 10.04.1012.

Clark, A. (1998b). 'Time and Mind', *Journal of Philosophy*: XCV: 7, pp.354-376

Clark, A. (1999). 'Embodiment: From Fish to Fantasy', URL: http://www.philosophy.ed.ac.uk/people/clark/pubs/TICSEmbodiment.pdf

Clark, A. (2001). *Mindware: An Introduction to the Philosophy of Cognitive Science*. Oxford: Oxford University Press.

Clark, A. (2003a). *Natural-Born Cyborgs: Minds, Technologies and the Future of Human Intelligence*. Oxford: Oxford University Press.

Clark, A. (2003b). 'Memento's Revenge: The Extended Mind, Extended.' URL: http://www.philosophy.ed.ac.uk/people/clark/pubs/Mementosrevenge2.pdf

Clark, A. (2008). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford: Oxford University Press.

Clark, A. & Grush, R. (1999). 'Towards a Cognitive Robotics'. *Adaptive Behavior*, 7(1), pp. 5-16

Cole, M. & Scribner, S. (1978*), Introduction to: Mind in Society: The Development of Higher Psychological Processes*. Cambridge, Massachusetts: Harvard University Press, pp.1-14.

Cornforth, M. (1952). *Dialectical Materialism.* Lawrence & Wishart: London.

Corning, Peter A. (*2002*), 'The Re-Emergence of "Emergence": A Venerable Concept in Search of a Theory', *Complexity* 7(6): pp. 18-30.

Dahlbom, B. (1993). 'Mind is Artificial'. In B. Dahlbom (ed.), Dennett and His Critics. Cambridge: Blackwell.

Damasio, A.R. (1994). *Descartes' Error: Emotion, Reason and the Human Brain*, New York: G.P. Putnam's Sons. Quill Paperback, 2000.

Damasio, A.R. (2000).*The Feeling of What Happens*. London: Vintage.

Damasio, A.R. (2003). *Looking for Spinoza.* London: Vintage.

Danto, A. (1973). *Analytical Philosophy of Action*. Cambridge: Cambridge University Press.

Dartnall, T. (2004). 'We Have Always Been . . . Cyborgs.' Metascience 13 (2): pp.139-273.

Dauve, G. & Martin, F. (1974). *The Eclipse and Re-Emergence of the Communist Movement*. London: Antagonism Press.

Debord, G. (1967). *Society of the Spectacle*. Reprinted 1983, Black & Red: Detroit, Michigan.

De Jaegher, H. & Di Paolo, E. (2007). 'Participatory Sense-Making: An enactive approach to social cognition'. *Phenomenology and the Cognitive Sciences*, 6(4), 485-507.

Dennett, D. (1981a). 'True Believers: The Intentional Strategy and Why it Works'. Reprinted in *Mind Design II*, Haugeland, J. (ed.) Cambridge, Massachusetts: MIT Press.

Dennett, D. (1981b) 'Where am I?' in D. Dennett, *Brainstorms: Philosophical Essays on Mind and Psychology,* pp. 310-323. Cambridge, Massachusetts: MIT Press.

Dennett, D. (1984). 'Cognitive Wheels: The Frame Problem of AI' in: C. Hookway (ed.) *Minds, Machines and Evolution: Philosophical Studies*, Cambridge: Cambridge University Press, pp. 129-51.

Dennett, D. (1991). *Consciousness Explained*. Boston: Little, Brown. Penguin Edition, 1993.

Dennett, D. (1993). 'Review of F. Varela, E. Thompson and E. Rosch, 'The Embodied Mind',' *American Journal of Psychology*, 106: pp. 121-6.

Dennett, D. (2003). *Freedom Evolves*. New York: Penguin.

Destefano, M., Lindstedt, J.K. & Gray, W.D. (2011). 'Use of Complementary Actions Decreases with Expertise' in L. Carlson, C. Hoelscher, & T.F. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2709-2714). Austin, TX: Cognitive Science Society.

Dreyfus, H. L. (1965). *Alchemy and Artificial Intelligence*, Santa Monica California: Rand Corp.

Dreyfus, H.L. (1972). *What Computers Can't Do – A Critique of Artificial Reason*, New York: Harper & Row.

Dreyfus, H.L. & Dreyfus, S. (2002). 'From Socrates to Expert Systems', *Philosophy* 24-1.

Di Giovanni, G. (ed. 1990). *Essays On Hegel's Logic*, New York: SUNY Press.

Di Paolo, E. (2005). 'Autopoiesis, Adaptivity, Teleology, Agency', *Phenomenology and the Cognitive Sciences*. 4: 429-452.

Di Paolo, E., Rhode, M., & De Jaegher, H. (2010) 'Horizons for the Enactive Mind: Values, Social Interaction, and Play.' in J. Stewart, O. Gapenne & E. Di Paolo (Eds.), Enaction: Towards a New Paradigm for Cognitive Science. Cambridge, MA: MIT Press.

Dunayevskaya, R. (1973). *Philosophy and Revolution*. New York: Delacorte Press.

Eldred, M. (2010). 'Capital and Technology: Marx and Heidegger'. *Left Curve No. 24*, May 2000, Oakland: California. Version 3.0. March 2010.

Eliasmith, C. (1996). 'The Third Contender: a Critical Examination of the Dynamicist Theory of Cognition'. *Philosophical Psychology*. Vol.9, No. 4 pp.441-463.

Eliasmith, C. (2001). 'Attractive and In-Discrete: A Critique of Two Putative Virtues of the Dynamicist Theory of Mind'. *Minds and Machines* 11: 417-426.

Engels, F.E. (1883). *The Dialectics of Nature*. London: Lawrence & Wishart. 1940.

Engels, F.E. (1887). *Anti-Duhring*. Moscow: Foreign Languages Publishing House. 1959.

Faith, J. (2000). 'Emergent Representations: Dialectical Materialism and the Philosophy of Mind', PhD Thesis, University of Sussex.

Fann, K.T. (2002). 'Beyond Marx and Wittgenstein (a Confession of a Wittgensteinian Marxist Turned Taoist)', in G. N. Kitching & Nigel Pleasants (eds.), Marx and Wittgenstein: Knowledge, Morality and Politics. Routledge.

Feuerbach, L. (1841). 'The Essence of Christianity', reprinted in Hanfi, Z. (ed. 1972*). The Fiery Brook: Selected Writings of Ludwig Feuerbach.* New York: Anchor Books, pp. 97-133.

Findlay, J.N. (1977). 'Analysis of the Text' in *Phenomenology of Spirit*, trans. by A.V. Miller, Oxford: Oxford University Press, pp. 495-591.

Fischer, E. (1968). *Marx in His Own Words*. Pelican edition, 1973.

Fodor, J.A. (1983). *The Modularity of Mind*, Cambridge Massachusetts: MIT Press.

Fodor, J.A. (2009) 'Where is my Mind?' *London Review of Books* 38 (3), pp. 13-15

Fodor, J.A., & Pylyshyn, Z.W. (1988). 'Connectionism and Cognitive Architecture: A Critical Analysis' in Haugeland, J. (ed. 1997). *Mind Design II*, Cambridge, Massachusetts: MIT Press, pp. 309-350.

Froese, T. (2007), 'On the role of AI in the Ongoing Paradigm Shift within the Cognitive Sciences', in: M. Lungarella et al. (eds.), *50 Years of AI*, Berlin, Germany: Springer Verlag, pp. 63-75

Froese, T. & Di Paolo, E. A. (2009), 'Sociality and the life-mind continuity thesis', Phenomenology and the Cognitive Sciences, 8(4), pp. 439-463

Fromm, E. (1966). *Marx's Concept of Man*. Continuum edition, 2003.

Gallagher, S. (2005). *How the Body Shapes the Mind*. Oxford: Oxford University Press.

Gallese, V. (2001). 'The Shared Manifold Hypothesis: From Mirror Neurons to Empathy' in *Between Ourselves – Second Person Issues in the Study of Consciousness*, (ed. E. Thompson) Journal of Consciousness Studies 8 (5-7):33-50.

Gallese, V. (2003). 'The Manifold Nature of Interpersonal Relations: the Quest for a Common Mechanism.' *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, Vol. 358, No. 1431, pp. 517-528

Gallese, V., & Lakoff, G. (2005). 'The Brain's Concepts: The Role of the Sensory-Motor System in Reason and Language'. *Cognitive Neuropsychology*, 22: 455-479.

Gardner, H. (1985). *The Mind's New Science: A History of the Cognitive Revolution.* New York: Basic Books.

Gibson, J.G. (1979), *The Ecological Approach to Visual Perception*, Boston: Houghton Mifflin.

Goldman, A.L. (1970). *A Theory of Human Action*. Eglewood Cliffs: Prentice Hall.

Goldman, A., & de Vignemont, F. (2009) 'Is Social Cognition Embodied?' *Trends in Cognitive Science*, 13:4. pp.154-159.

Gorz, A. (1983). *Paths to Paradise: On the Liberation From Work*. Paris: *Galilée.*

Hanfi, Z. (ed. 1972). *The Fiery Brook: Selected Writings of Ludwig Feuerbach.* New York: Anchor Books.

Hanna, R. (1996). 'From an Ontological Point of View: Hegel's Critique of the Common Logic', in *The Hegel Myths and Legends*, Stewart, J., Illinois: Northwestern University Press.

Hardcastle, V.G. (1999). 'It's O.K. to Be Complicated: The Case of Emotion'. *Journal of Consciousness Studies*, 6 (11-12), 237, 249.

Harnad, S. (1990). 'The Symbol Grounding Problem'. *Physica* D 42: 335-346.

Harnad, S. (2003). 'Can a Machine Be Conscious? How? *Journal of Consciousness Studies* 10(4-5): 69-75.

Hartnack, J. (1998). *An Introduction to Hegel's Logic*, Indianapolis: Hackett Publishing.

Haugeland, J. (1985). *Artificial Intelligence: The Very Idea*, Cambridge Massachusetts: MIT Press.

Haugeland, J. (1996). 'What is Mind Design?' in *Mind Design II*, (ed. J. Haugeland), Cambridge, Massachusetts: MIT Press

Haugeland, J. (ed. 1997). *Mind Design II*, Cambridge, Massachusetts: MIT Press.

Hegel, G.W.F. (1802). *System of Ethical Life*. URL: http://www.marxists.org/reference/ archive/hegel/works/se/ch01.htm Accessed: 10 April, 2012.

Hegel, G.W.F. (1805-6a). *Jena Lectures – Part 1 – Conclusion*. URL: http://www.marxists.org/ reference/archive/hegel/works/jl/ch01c.htm. Accessed: 10 April, 2012.

Hegel, G.W.F. (1805-6b). *Jena Lectures – Part 2 – Actual Spirit*. URL: http://www.marxists.org/reference/archive/hegel/works/jl/ch02a.htm. Accessed: 10 April, 2012.

Hegel, G.W.F. (1807). *Phenomenology of Spirit*, trans. by A.V. Miller, Oxford: Oxford University Press, 1977

Hegel, G.W.F (1812-1816). *Science of Logic*, trans. By A.V. Miller, London: Routledge, 2002.

Hegel, G.W.F. (1817a). *Hegel's Logic (Part One of the Encyclopaedia of Philosophical Sciences)* trans. by W. Wallace, Oxford: Oxford University Press, 1975.

Hegel, G.W.F. (1817b). *Philosophy of Nature* (*Part Two of the Encyclopedia of the Philosophical Sciences*), ed. E. Behler, trans by S. Taubeneck, New York: Continuum, 1990

Hegel, G.W.F. (1817c). *Philosophy of Mind* (*Part Three of the Encyclopedia of the Philosophical Sciences*), trans by W. Wallace, New York: Cosimo, 2008.

Hegel, G.W.F., *Philosophy of Right*, (1821), trans. by S.W. Dyde, New York: Prometheus Books, 1996.

Hegel, G.W.F. (1822/31), *Philosophy of History*, trans. by J.Sibree, New York: Dover Publications, 1956.

Hegel, G.W.F. (1835). *Hegel's Aesthetics Vol.1*, trans. by T.M. Knox, Oxford: Oxford University Press, 1998.

Heidegger, M. (1927). *Being and Time.* Oxford: Blackwell edition (trans. J. Macquarrie & E. Robinson), reprinted 1987.

Heidegger, M. (1977). *The Question Concerning Technology, and Other Essays*. New York: Harper & Row.

Heine, H. (1832). *Religion and Philosophy in Germany*. Albany: State University of New York Press, 1986.

Honderich, T. (1988). *Mind and Brain – A Theory of Determinism – Volume 1*. Oxford: Clarendon Press.

Holloway, J. (2002). *Change the World Without Taking Power*. London: Pluto Press.

Humphries, C. (1971). 'Two Marx's or One?', *International Socialism* 1: 49, pp. 32-33.

Hurley, S. (2001). 'Perception and Action: Alternative Views', *Synthese* 129: pp. 3-40.

Hutchison, W. D. *et al.* (1999). 'Pain-related Neurons in the Human Cingulated Cortex'. *Nature Neuroscience.* 2 (5), pp. 403 – 405.

Ingold, T. (1987) *The Appropriation of Nature: Essays on Human Ecology and Social Relations*. Iowa: University of Iowa Press.

Inwood, M.J. (1989). *Hegel – Selections*. London: Collier Macmillan.

Iverson, J.M.,& Thelen, E. (1999). 'Hand, Mouth and Brain, The Dynamic Emergence of Speech and Gesture', in *Reclaiming Cognition*, (Journal of Consciousness Studies, 6, No. 11-12), Nunez, R, & Freeman, W.J. (eds.), Thorverton, Exeter: Imprint Academic.

Jackendoff, R. (1987). *Consciousness and the Computational Mind*. Cambridge, Massachusetts: MIT Press.

Jha, S. (2010). 'Exploring the Horizon: Constructing Marx's Contribution on Praxis as Transcendence', *Critique*, 38:2, pp. 207-218.

Johnson, D.G. & Nissenbaum, H. (eds. 1995). *Computers, Ethics and Social Values*, Englewood Cliffs, NJ: Prentice Hall.

Kelso, J.A.S. (1995). *Dynamic Patterns: The Self Organization of Brain and Behavior*. Cambridge, Massachusetts: MIT Press. Paperback edition, 1997.

Kirsh, D. & Maglio, P. (1994). 'On Distinguishing Epistemic from Pragmatic Action', *Cognitive Science*, Volume 18, 4: pp. 513-549.

Kirsh, D. 1995. 'The Intelligent Use of Space'. *Artificial Intelligence* 73: pp.31-68.

Kirsh, D. (2006). 'Distributed Cognition – A Methodological Note', *Pragmatics and Cognition*, 14, 2: pp. 249-262.

Kirsh, D. (2010). 'Thinking with External Representations', *AI & Society* 25, pp.441-454.

Kitching, G.N. (1988). *Karl Marx and the Philosophy of Praxis.* London: Routledge.

Kitching, G.N. & Pleasants, L. (eds. 2002). *Marx and Wittgenstein: Knowledge, Morality and Politics*. London: Routledge.

Kiverstein, J. & Wheeler, M. (eds. 2012). *Heidegger and Cognitive Science*. Palgrave, Macmillan.

Kripke, S.A. (*1982*). *Wittgenstein on Rules and Private Language*. Oxford: Basil Blackwell.

Kuhn, T.S. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Labriola, A. (1897). Socialism and Philosophy IV-VI. Online at: http://www.marxists.org/archive/labriola/works/al04.htm

Lakoff, G. & Johnson, M. (1980) *.Metaphors We Live By*. Chicago: University of Chicago Press.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind.* Chicago: University of Chicago Press.

Lakoff, G. (1988). 'Cognitive Semantics', in *Meaning and Mental Representations*, ed. Umberto Eco et. al. Bloomington: Indiana University Press: pp.119-154.

Lakoff, G. & Johnson, M. (1999). *Philosophy In The Flesh: The Embodied Mind and Its Challenge to Western Thought*. New York: Basic Books

Lenin, V.I. (1908). *Materialism and Empirio-Criticism*. 1977 edition, Progress Publishers: USSR.

Levine, N. (1975). *The Tragic Deception: Marx Contra Engels*. Oxford: Clio Books.

Levins, R. & Lewontin, R. (1985). *The Dialectical Biologist*. Cambridge, Massachusetts: Harvard University Press.

Lewontin, R. (1983). 'The Organism as the Subject and Object of Evolution'. *Scientia* 118: pp. 63-82.

Liebich, A. (ed. 1979). *Selected Writings of August Cieszkowski*. Cambridge: Cambridge University Press.

Lukacs, G. (1975). *The Young Hegel* (orig. pub 1947*)*, London: Merlin Press.

Lumsden, S. (2008), 'The Rise of the Non-Metaphysical Hegel', *Philosophy Compass*, 3: pp. 51–65.

Lycan, W.G., (ed.1999). *Mind and Cognition: An Anthology*, 2nd Edition. Malden, Massachusetts: Blackwell.

Mackenzie, A. (2004). Review of Clark, A. (2003a) in 'We Have Always Been . . . Cyborgs.' Metascience 13 (2): pp.139-273.

Maglio, P.P. & Kirsh, D. (1996). 'Epistemic Action Increases with Skill'. *Proceedings of the Eighteenth annual conference of the cognitive science society*, Mahwah, NJ: Erlbaum, 391–396.

Maglio, P., Matlock, T., Raphaely, D., Chernicky, B., & Kirsh D. (1999). 'Interactive skill in Scrabble'. *Proceedings of the Twenty-first Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum.

Maglio, P.P., Wenger, M.J. & Copeland, A.M. (2003). 'The Benefits of Epistemic Action Outweigh the Costs.' *Proceedings of the 25th annual conference of the Cognitive Science Society*, 752-757.

Maglio, P. M., Wenger, M. J., & Copeland, A. M. (2008). 'Evidence for the Role of Self-Priming in Epistemic Action: Expertise and the Effective use of Memory'. *Acta Psychologica,* 127, 72-88

Malatesta, E. (1891). *Anarchy*. Freedom Press edition, 1974.

Marcuse, H. (1968). *Reason and Revolution*, (orig. pub. 1941), London: Routledge.

Marcuse, H. (2005). *Heideggerian Marxism*. (eds. R. Wolin & J. Abromeit). Lincoln: University of Nebraska Press.

Marx, K. (1844a), 'Economic and Philosophic Manuscripts', reprinted in Mclellan, D. (ed.), *Karl Marx, Selected Writings*, Oxford: Oxford University Press. 1977, pp 75-111.

Marx, K. (1844b). 'On James Mill', reprinted in Mclellan, D. (ed.), *Karl Marx, Selected Writings*, Oxford: Oxford University Press. 1977, pp 114-123.

Marx, K. (1846). *The Poverty of Philosophy*. Foreign Languages Press edition: 1978.

Marx, K. & Engels, F. (1848) *The Communist Manifesto*, Oxford: Oxford University Press, 1998.

Marx, K. (1855). *Grundrisse – Foundations of the Critique of Political Economy*. Penguin edition 1973.

Marx, K. (1867). *Capital – A Critique of Political Economy, Volume 1*. Penguin edition, 1976.

Marx, K. (1873). 'Afterword to the Second German Edition of Capital', in Capital Vol.1, trans. Samuel *Moore* and Edward *Aveling* , Moscow: *Progress Publishers*, 1974. pp.10-15.

Marx, K. (1970a). *Thesis on Feuerbach*. Reprinted in The German Ideology *(ed. C.Arthur)*, London, Lawrence & Wishart, revised edition 1974.

Marx, K. and F. Engels (1970b). *The German Ideology* (ed. C. Arthur), London: Lawrence & Wishart: revised edition 1974.

Maturana, H. R. & Varela, F. J. (1980). *Autopoiesis and Cognition- The Realization of the Living*. (Boston studies in the philosophy of science; v.42) Boston: Kluwer.

Maturana, H. R. & Varela, F. J. (1987). *The Tree of knowledge: The Biological Roots of Human Understanding*. Boston: Shambhala Publications. Revised edition, 1998.

McCarthy, J. & Hayes, P. J. (1969). 'Some Philosophical Problems from the Standpoint of Artificial Intelligence', in Meltzer, B. and Michie, D. (eds.), *Machine Intelligence* 4, pp. 463-502. Edinburgh: Edinburgh University Press.

McLellan, D. (1969). *The Young Hegelians and Karl Marx*, London: Macmillan.

Mclellan, D. (ed. 1977). *Karl Marx, Selected Writings*, Oxford: Oxford University Press.

McLellan, D. (1979). *Marx's Grundrisse*, London: Granada Publishing Ltd.

Menary, R. (ed. 2010). *The Extended Mind*. Massachusetts: MIT Press.

Merleau-Ponty, M. (1942). *The Structure of Behaviour*. France: Presses Universitaires de France. Reprinted by Duquesne University Press, Seventh Printing, 2002.

Merleau-Ponty, M. (1945). *The Phenomenology of Perception*. Paris: Gallimard. Reprinted by Routledge & Kegan Paul, 1962.

Merleau-Ponty, M. (1947). *Humanism and Terror – An Essay on the Communist Problem.* Paris: Gallimard. Reprinted by Beacon Press: Boston, 1969.

Merleau-Ponty, M. (1955). *Adventures of the Dialectic.* Paris: Gallimard. Reprinted by Northwestern University Press, 1973.

Mészáros, I. (1970) *Marx's Theory of Alienation*. London: Merlin Press.

Mill, J.S. (1889). An Examination of Sir William Hamilton's Philosophy. London and New York: Longman's, Green & Co.

Miller, J. (1976). 'Merleau-Ponty's Marxism: Between Phenomenology and the Hegelian Absolute', *History and Theory*, Vol.15: 2, pp.109-132.

Minsky, M. (1986) *The Society of Mind*. New York: Simon and Schuster

Morrison, I. et al. (2004) 'Vicarious Responses to Pain in Anterior Cingulate Cortex: Is Empathy a Multisensory Issue?' *Cognitive, Affective, & Behavioral Neuroscience*, 2004, 4 (2), pp. 270-278.

Moyal-Sharrock, D. (2009). 'Wittgenstein and the Memory Debate.' *New Ideas in Psychology* Special Issue: 'Mind, Meaning and Language: Wittgenstein's Relevance for Psychology' 27: pp. 213-27.

Mueller, G.E. (1996). 'The Hegel Legend of 'Thesis-Antithesis-Synthesis' in Stewart, J. (1996). *The Hegel Myths and Legends*, Illinois: Northwestern University Press.

Newell, A, & Simon N. (1985). 'Computer Science as Empirical Inquiry: Symbols and Search', in *Mind Design II*, Haugeland, J. (ed.), Cambridge, Massachusetts: MIT Press.

Newell, A. (1995). 'The Serial Imperative' in Baumgartner, P., & Payr, S. (eds. 1995). *Speaking Minds: Interviews with Twenty Eminent Cognitive Scientists.* Princeton: Princeton University Press, pp. 145-155.

Noë, A. (2002). 'Is the Visual World a Grand Illusion', *Journal of Consciousness Studies*. Volume 9, No 5/6, pp. 1-12.

Noë, A. (2004) *Action in Perception. The Enactive Approach to Perception*. Cambridge, Massachusetts: MIT Press.

Norman, R. (1976). *Hegel's Phenomenology – A Philosophical Introduction*, Sussex: Sussex University Press.

Novack, G. (2002). *Marxist Writings on History and Philosophy*. Australia: Resistance Books.

Nunez, R, & Freeman, W.J. (eds. 1999), *Reclaiming Cognition*, (Journal of Consciousness Studies, 6, No. 11-12), Thorverton, Exeter: Imprint Academic.

O'Brien, G. (1998). Review of Clark, A (1997). *Metascience* 7, pp. 78-83.

Ollman, B. (1971). *Alienation – Marx's Conception of Man in Capitalist Society*. Cambridge: Cambridge University Press. Second Edition, 1976.

O'Regan, J.K., & Noë, A. (2001). 'A Sensorimotor Account of Vision and Visual Consciousness'. *Behavioral and Brain Sciences, 24*(5), pp. 939-1031.

Pannekoek, A. (1937). 'Society and Mind in Marxian Philosophy', *Science in Society*, vol. no.4, pp. 445-453.

Patten, A. (1999). *Hegel's Idea of Freedom*, Oxford: Oxford University Press.

Petrovic, G. (1991). 'Reification', in Bottomore, T. (ed. 1991). *A Dictionary of Marxist Thought*. Oxford: Blackwell.

Pippin, R. (1996). 'Hegel's Metaphysics and the Problem of Contradiction', in *The Hegel Myths and Legends*, Stewart, J., Illinois: Northwestern University Press.

Popper, K.R. (1945). *The Open Society and its Enemies, Vol. 2 Hegel and Marx*, London: Routledge. Revised edition, 1966.

Potter, G. (2008). 'Humanism and Terror: Merleau-Ponty's Marxism', in *Cultural Logic*, 2008. [online] Available at: http://clogic.eserver.org/2008/2008.html. [Accessed 26 January 2012]

Prinz, J.J. (2005). 'Are Emotions Feelings?' *Journal of Consciousness Studies* 12 (8-10):9-25.

Putnam, H. (1969). 'Psychological Predicates', in W. H. Capitan & D. D. Merrill (eds.), Art, Mind, and Religion. University of Pittsburgh Press.

Ramsey, W., S. Stitch, and J. Garon: 1990, 'Connectionism, Eliminativism, and the Future of Folk Psychology', in J. Tomberlin (ed.), *Philosophical Perspectives*, 4, Ridgeview, Atascadero, California. Reprinted in J. Haugeland (ed.) *Mind Design II*. pp. 351-376.

Rauch, L. (1983). *Hegel and the Human Spirit. A Translation of the Jena Lectures on the Philosophy of Spirit* (*with commentary)*. Detroit: Wayne State University Press.

Rees, J. (1998). *The Algebra of Revolution: The Dialectic and the Classical Marxist Tradition.* London: Routledge.

Robbins, P, & Aydede, M. (2009). *The Cambridge Handbook of Situated Cognition*, Cambridge: Cambridge University Press.

Roszak, T. (1969). *The Making of a Counter Culture*, New York: Doubleday & Co.

Roszak, T. (1986). *The Cult of Information*, Cambridge (Lutterworth Press).

Rowlands, M. (2010). 'Consciousness, Broadly Construed' in Menary, R. (ed.), *The Extended Mind*, Cambridge, Massachusetts: MIT Press.

Rubinstein, D. (1981). *Marx and Wittgenstein – Social Praxis and Social Explanation*, Routledge: London

Rupert, R, D. (2009). *Cognitive Systems and the Extended Mind.* Oxford: Oxford University Press.

Russell, B. (1912). *The Problems of Philosophy*. Oxford: Oxford University Press (Reprinted 1973).

Russell, B. (1914). *Our Knowledge of the External World*. London: George Allen & Unwin.

Russell, B. (1950). *Unpopular Essays*, London: George Allen & Unwin.

Russell, S. & Norvig, P. (1995) *Artificial Intelligence: A Modern Approach.* London: Prentice Hall.

Ryle, G. (1949). *The Concept of Mind.* London: Hutchinson & Company.

Sayers, S. (2003). 'Creative Activity and Alienation in Hegel and Marx', *Historical Materialism*, 11:1, pp.107-128.

Sayers, S. (2012). *Marx and Alienation.* Palgrave, Macmillan.

Searle, J.R. (1980). 'Minds, Brains, and Programs', *Behavioural and Brain Sciences* 3 (3): pp. 417-457

Simon, H.A. (1995). 'Machine as Mind', in K.M. Ford, C. Glymour, & P.J. Hayes (eds.), *Android Epistemology*, Cambridge, Massachusetts: MIT Press.

Smith, A. (1776). *The Wealth of Nations, Books 1-3*, Penguin edition, 1999.

Sohn-Rethel, A. (1978), *Intellectual and Manual Labour – A Critique of Epistemology*, London: Macmillan.

Steels, L. (1994). 'The Artificial Roots of Artificial Intelligence'. *Artificial Life* 1, (1-2): pp. 75-110.

Stepelevich, L. (ed. 1983). *The Young Hegelians*, Cambridge: Cambridge University Press.

Stern, R. (2002). *Hegel and the Phenomenology of Spirit*, London. Routledge.

Stewart, J. (1996). *The Hegel Myths and Legends*, Illinois: Northwestern University Press.

Stillman, P.G. (ed. 1987). *Hegel's Philosophy of Spirit*, New York: SUNY Press.

Suchman, L.A. (1987). *Plans and Situated Actions: The Problem of Human Machine Communication*. Cambridge: Cambridge University Press.

Taylor, C.T. (1975). *Hegel*. Cambridge: Cambridge University Press.

Thelen, E., and L. Smith (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, Massachusetts: MIT Press.

Thompson, E. (ed. 2001). Between Ourselves: Second-person Issues in the Study of Consciousness. *Journal of Consciousness Studies* 8 (5-7). Thorverton: Imprint Academic.

Thompson, E. (2001). 'Empathy and Consciousness' in Between Ourselves: Second-person Issues in the Study of Consciousness. *Journal of Consciousness Studies* 8 (5-7). Thorverton: Imprint Academic: pp.1-32.

Thompson, E. (2004). 'Life and Mind: From Autopoiesis to Neurophenomenology. A Tribute to Francisco Varela', Phenomenology and the Cognitive Sciences, 3(4), pp. 381-398.

Thompson, E. (2007). *Mind in Life*. Cambridge, Massachusetts: Harvard University Press.

Torrance, S. (1999). 'Real-world Embedding and Traditional Artificial Intelligence', in *Perspectives on Cognitive Science*, Vol.2, Wiles, J, & Dartnall, T. (eds.), Stamford , CT: Ablex Publishing Corporation.

Torrance, S. (2005). 'In Search of the Enactive: Introduction to Special Issue on Enactive Experience', *Phenomenology and the Cognitive Sciences,* 4(4) December 2005, pp. 357-368.

Torrance, S. & Froese, T. (2011). 'An Inter-Enactive Approach to Agency: Participatory Sense-Making, Dynamics and Sociality', *Humana.Mente,* 15, pp. 21-53.

Toth, J.P. & Hunt, R.R. (1999). 'Not One Versus Many, but Zero Versus Any: Structure and Function in the Context of the Multiple Memory-systems Debate.' in J.K. Foster and M. Jelicic (eds.), *Memory: System, process, or function?* (pp. 232-272), Oxford: Oxford University Press

Toulmin, S. (1978). 'The Mozart of Psychology'. *New York Review of Books*, 25(14), pp. 51–57. Accessed at www.nybooks.com (no individual page refs. available)

Turing, A.M. (1950). 'Computing Machinery and Intelligence', reprinted in Haugeland, J. (ed. 1997). *Mind Design II*, Cambridge, Massachusetts: MIT Press, pp. 29-56.

Vaneigem, R. (1967). *The Revolution of Everyday Life.* 1983 Edition, London: Aldgate Press.

van Gelder, T. (1995). `What might cognition be if not computation?' *Journal of Philosophy 92*: pp. 345-381. Reprinted in W.G. Lycan (ed.) *Mind and Cognition*, pp. 170-191.

van Gelder, T. (1996). 'Dynamics and cognition', in J. Haugeland (ed.) *Mind Design II*. pp. 421-450.

van Gelder, T. & Port, R.F. (1995). 'It's About Time: An Overview of the Dynamical Approach to Cognition', in Mind as Motion: Explorations in the Dynamics of Cognition (eds. Van Gelder, T. & Port, R.F.) Cambridge, Massachusetts: MIT Press, pp.1-43.

van Heijenoort, J. (1948) 'Friedrich Engels and Mathematics', in *Selected Essays* (Napoli: Bibliopolis, 1985), pp. 123-151.

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, Massachusetts: MIT Press. Paperback edition, 1993.

Varela, F.J., Shear J., (eds. 1999). *The View from Within: First-Person Methodologies in the Study of Consciousness*. Special Issue of Journal of Consciousness Studies, 6 (2-3)

Velli, M. (1972). *The Seizure of State Power*. London: Black & Red. (1992 Edition)

Veresov, N. (2005). 'Marxist and non-Marxist Aspects of the Cultural-Historical Psychology of L.S. Vygotsky', *Outlines*, 1, pp. 31- 49.

Vygotsky, L. (1930). 'The Instrumental Method in Psychology', in R. van der Veer (Ed.), *The Collected Works of L.S. Vygotsky: Volume 3: Problems of the Theory and History of Psychology*. New York: Plenum Press.

Vygotsky, L. (1934). *Thought and Language*. Revised edition (ed. & trans. A. Kozulin), Massachusetts: MIT Press. 1986.

Vygotsky, L (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, Massachusetts: Harvard University Press.

Walmsley, J. (2003). *Dynamical Systems, Emergence and Explanation*. URL: http://www.lehigh.edu/~interact/isi2003/ISI%202003%20Web%20Papers/ WalmsleyCopenhagenPaperv.4.pdf

Weiss, F.G. (1974). *Hegel: The Essential Writings*. New York: Harper &Row.

Weizenbaum, J. (1972). 'On the Impact of the Computer on Society,' *Science*, Vol. 176, No. 12, pp. 609-614. Reprinted in D. Johnson and H. Nissenbaum, eds. *Computers, Ethics & Social Values*. Englewood Cliffs, NJ: Prentice Hall, 1995, pp. 549-559

Weizenbaum, J. (1976). *Computer Power and Human Reason.* San Francisco: Freeman.

Wendling, A. (2009). *Karl Marx on Technology and Alienation*. London: Palgrave Macmillan. First paperback edition: 2011.

Westphal, K.R. (2003). *Hegel's Epistemology*. Indianapolis: Hackett Publishing.

Wheeler, M. (2005). *Reconstructing the Cognitive World: The Next Step*, Massachusetts: MIT Press.

Wiles, J, & Dartnall, T. (ed. 1999). *Perspectives on Cognitive Science*, Vol.2., Stamford , CT: Ablex Publishing Corporation.

Williams, R.R. (1987). 'Hegel's Concept of Geist', in *Hegel's Philosophy of Spirit*. Stillman, P.G. (ed.), New York: SUNY Press.

Wilson, F.R. (1999). *The Hand : How its use Shapes the Brain, Language and Human Culture*. New York: Random House.

Wilson, M. (2002). 'Six Views of Embodied Cognition,' *Psychonomic Bulletin and Review*, 9, pp. 625-636.

Wittgenstein, L. (1953). *Philosophical Investigations*, Oxford, Basil Blackwell Ltd.

Wolin, R. (1993). *The Heidegger Controversy*. Cambridge, Massachusetts: MIT Press.

Wollin, R. & Abromeit, J. (eds. 2005) *Heideggerian Marxism. Herbert Marcuse*. Lincoln: University of Nebraska Press.

Wood, D., Bruner, J.S. et al. (1976), 'The Role of Tutoring in Problem Solving', *Journal of Child Psychology and Psychiatry and Allied Disciplines*, vol. 17 (2), pp.89-100

Zahavi, D. (2001). 'Beyond Empathy: Phenomenological Approaches to Intersubjectivity', in Thompson, E. (ed. 2001). Between Ourselves: Second-person Issues in the Study of Consciousness. *Journal of Consciousness Studies* 8 (5-7). Thorverton: Imprint Academic, pp.151-167.

Ziemke, T. (2003). 'What's That Thing Called Embodiment?' in: Alterman & Kirsh (eds.) Proceedings of the 25th Annual Conference of the Cognitive Science Society (pp. 1134-1139). Mahwah, NJ: Lawrence Erlbaum.