THE FLESH IS WEAK

BRENDAN WILSON

University of Tokyo

Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought, by George Lakoff and Mark Johnson, Basic Books, New York, 1999, xiv+624pp.

Keywords: cognitive science, mind/brain identity, metaphor, conceptualisation, embodiment

1. Introduction

Some very common verbs in English have irregular past tense forms, and children typically acquire these irregular past tenses in a rather irregular way. At first they get them right, pairing *go* with *went*, *take* with *took* and so on. Then they make systematic mistakes, pairing *go* with *goed*, *take* with *taked*. Then—eventually—they get them right again. Sometimes, in between stage 2 (getting them wrong) and stage 3 (getting them right again), they get them very wrong indeed, pairing *go* with *wented*, *take* with *tooked* etc. This is a perplexing sequence of events, and we naturally want an explanation.

A theory in the cognitive style takes the view that this complicated overt behaviour is to be explained by developments in the learner's mind. A cognitive theory might claim, for example, that the learner formulates a rule (unconsciously) which states that past tenses are formed by adding *-ed*. Before the learner forms this rule, he or she simply remembers *go-went* as a pair, and so gets the past tense right. Once the rule is in place, the learner applies it to all verbs indiscriminately, producing *goed*. Sometime later, the stored *go-went* pair comes back into play, occasionally combining with the rule to produce *wented*. Finally, the stored *go-went* pair stabilises as a full exception to the 'Add *-ed*' rule.

As this example shows, the cognitive revolution in linguistics and psychology was a revolution against behaviourism. Cognitive science is *cognitive* because it claims that overt behaviour cannot be explained without invoking mental entities and processes. And it's *science* (most of its practioners suppose) because its explanations are causal. The mental representation of the rule 'Add -ed', or of the *go-went* pair, makes a *causal* difference to the physical process of uttering the word *goed* or *wented*.

But of course, even the most committed cognitive scientist feels a bit uneasy about this. *How* exactly is a mental representation supposed to exert its causal power? Being mental, it has none of the physical, or chemical, or electrical properties by which other causes work. It isn't solid, nor yet liquid, nor a gas. It doesn't turn litmus paper blue, or for that matter red. It neither conducts nor resists the flow of electrons. *How* is it supposed to work on entities and processes which do have these useful properties? Trying to understand how this alleged causal influence might operate is, as T H Huxley said, like trying to understand how one might hit a nominative case with a stick.¹

Of course, there is a response to this problem—some version or other of the mind-brain identity theory. If the mental representation of the rule 'Add -ed' is identical with (is the very same thing as) some brain state or process, then *qua* brain state, it has the necessary material properties to be a perfectly respectable cause. Chomsky (2000: 25) says, 'We assume, essentially on faith, that there is some kind of description [of representations] in terms of atoms and molecules.' Steven Pinker (1994: 78) says that mind/brain identification is 'as fundamental to cognitive science as the cell doctrine is to biology and plate tectonics is to geology,' and Lakoff and Johnson (1999: 110) describe it as 'a common paradigm that most cognitive scientists share.' It's important to realise how universal this resort to mind/brain identity is, because it is obviously doomed.²

¹ Jackendoff (1990), for example, takes this to mean that any attempt to explain physical effects by reference to phenomenological causation 'amounts essentially to an appeal to magic,' and accepts the 'ugly consequence [that] *consciousness is not good for anything* (p. 26)'

² Ronald Langacker's connectionist model is another version of mind-brain identity. See Langacker (1990: 282f), and for a short critique, Wilson (1998: 126–8).

The point, remember, is to calm our disquiet about ascribing causal powers to cognitive entities and processes. But how are we to discover the alleged identity between cognitive and neural items *without* ascribing causal powers to the former? Consider the following well-worn example. We begin with a set of beliefs about the causal role of 'temperature'. Temperature can be raised by proximity to something hot, or by friction, or by certain chemical reactions. Heating things up makes them glow red, or burst into flame, or melt. Cooling things down turns liquids into solids, preserves food, and so on. All this constitutes the known causal role of 'temperature'. We then develop a theory of molecules, and it occurs to us that all these causes and effects would also be causes and effects of the average kinetic energy of the molecules. Since average kinetic energy does everything that temperature can do, and a lot more besides (like predicting microwave heating), we *identify* temperature with the average kinetic energy of molecules. We say, as Bacon did, not that molecular motion generates heat, but rather that 'Heat itself, its *quid ipsum*, is Motion and nothing else.'

In this way, discovering an identity relies precisely on knowing the causal powers of the terms identified, independently of the identification, and then realising that the causal role of one subsumes that of the other. So to discover an identity between mental events and neural ones, we would have to independently know the causal role of each, then realise that the causal role of the neural subsumes that of the mental. A mind/brain identity theory, therefore, takes it for granted that mental events have a causal role independent of the identification. In short, it takes for granted the very thing it was supposed to render intelligible. It answers the question—how could a cognitive entity be a cause?—by begging it.

The moral is that mind/brain identity theories, of whatever type, cannot soothe our post-Cartesian anxieties about non-physical causes. If we are unhappy about attributing a causal role to mental entities and processes, we cannot be in any way reassured by an alleged identity with brain entities and processes, since the identity depends on precisely the thing we are unhappy about.³

³ The eliminative materialism of Paul and Patricia Churchland neatly avoids this

So cognitive science has a problem: how can an entity or process with no physical properties be a cause? In fact, it has two problems. In addition to this metaphysical problem, there is an epistemological problem, which is at least as serious: how could the alleged causal influence of the cognitive entity or process be observed? If cognitive theories offer causal explanations, and are genuinely scientific, then there ought to be some way we could at least hope to observe these causal processes in operation. But of course, this is pie in the sky. They could only be observed if we had already established the mind/brain identity which depends on them.

2. Embodiment

It follows that cognitive science has every motivation—as we all do—to find something new to say about the mind/body problem, and the ambitious aim of Lakoff and Johnson's *Philosophy in the Flesh* is exactly this, to reconfigure the relationship of mind and body. The attempt, if successful, or even if merely interesting, certainly would represent the 'Challenge to Western Thought' trumpeted in their subtitle, since the two problems—metaphysical and epistemological—have been high on the agenda of Western Thought since Descartes.

Lakoff and Johnson organise their Challenge around three claims—that mind is essentially embodied, that abstract concepts are mostly metaphorical, and that most thought is unconscious. They believe that 'when taken together and considered in detail, these three findings from the science of the mind ... require our culture to abandon some of its deepest philosophical assumptions (p. 3).'

Drawing on the work of Eve Sweetser (1991), they point out that we often speak about ideas and the mind metaphorically. The mind can be spoken of as something which moves, and a line of thought as the path it follows. To think clearly is to think *straight*. A second analogy for thinking is perception, as when we *see* what someone means. A third analogy regards ideas as objects, which we manipulate in thinking. We can *grasp* an implication, for example, or *bounce* an idea off

problem, but a cognitive theory, which regards mental entities as causes, will naturally be unhappy to see them eliminated.

someone. Finally, thinking can also be regarded as ingestion, ideas as food, and the mind, correspondingly, as the digestive system. Readers of Lakoff and Johnson will be aware of the dangers of *swallowing half-baked* ideas.

So far, so uncontroversial. Nobody doubts that we talk about thinking, ideas and minds in these metaphorical ways. Wittgenstein, for example, summarises all this as 'the inner/outer analogy'.⁴ The same point is perfectly explicit in Reid.⁵ But according to Lakoff and Johnson, these and other common metaphors 'define our conceptualization of what ideas are and what thinking is (p. 248).' And this is problematic, since the metaphors have 'entailments' which are often singly unacceptable, and jointly incompatible. Insofar as ideas are food, we can *spoon-feed* them to someone. Can we therefore pick them up with chopsticks? And is it really polite to *bounce* them off someone, after we have *chewed them over*?

This is no joke, however, but the crux of the mind-body problem, according to Lakoff and Johnson. Any theory of mind, they claim, must 'pick a consistent subset of the entailments of these metaphors ... necessarily [leaving] behind other entailments, inconsistent with these, that are also 'intuitive' (p. 248).' Different theories of the mind, they believe, simply adopt different metaphors from ordinary language and work out their entailments. Apart from creating disputes between theories—disputes which are essentially metaphorical in character—these metaphors from ordinary language ignore modern research (which conflicts with them). And worst of all, since our metaphors are more or less united in implying that the mind is something distinct from the body, they hide from us 'the most central property of mind, its embodied character (p. 266).'

So the position is this: we can't avoid using some or other of these

⁴ See for example, Wittgenstein (1991 reprint: II.xi p. 196) and especially Wittgenstein (1958: *Blue Book*, passim).

⁵ Thomas Reid (1764: Ch. 7): 'The condition of mankind, therefore, affords good reason to apprehend that their language, and their common notions concerning the mind and its operations, will be analogical, and derived from the objects of sense; and that these analogies will be apt to impose upon philosophers, as well as upon the vulgar.'

The same theme reappears in Reid (1785: see for example Essay 4, Chapters 1 and 2).

metaphors. 'We cannot comprehend or reason about the mind without such metaphors (p. 266)': 'there can be no such thing as a non-metaphorical theory of mind (p. 409).' But using them creates philosophical disputes, prejudices us against modern discoveries, and falsifies the essential nature of the mind.

What are we supposed to do in this catch-22 situation? Happily, the inescapable influence of the metaphors we use for the mind has not hidden its 'most central property' from Lakoff and Johnson. Perhaps we too can come to see that the mind is embodied. Before we try to get clear about what the embodiment of mind is supposed to mean, however, let's be clear that, so far, we have nothing that will help cognitive explanation. Our question was: how can a mental entity be a cause? If mental entities, unlike the humble billiard ball, can only be understood metaphorically, that simply places their ability to function as causes under a further cloud.

What does it mean, then, to say that the mind is essentially embodied? In fact, it means quite different things in different parts of Lakoff and Johnson's book. Here's one explanation: 'The public nature of linguistic meaning and those aspects of meaning that are universal across cultures arise from the commonalities of our bodies and our bodily and social experience in the world (p. 462).' This is taken to oppose Frege's sharp distinction between the psychological associations of a word and its meaning, since this distinction is supposed to commit us to linguistic meaning as something completely divorced from human thought. 'Arising from the commonalities of our bodies' seems to mean that we form the concepts we do because we have the bodies we have. Thus, 'words ... express concepts, which reside in human minds and which ... get their meaning via their embodiment. Each of us, from childhood on, forms conceptual categories through our embodied experiences (p. 442).'

Talking of an 'embodied concept' then, as Lakoff and Johnson very frequently do, has the potential to mislead. They do not mean by this that the concept itself is embodied, present and locatable in the person's body somewhere. On the contrary, concepts 'reside in minds'. To say that a concept is embodied is to say that its meaning depends in some important way on the fact that the person in whose mind it resides, has a body: 'human concepts ... are crucially shaped by our bodies and brains, especially by our sensorimotor system (p. 22).'

Well, this sounds very much like good old-fashioned empiricism.

'*Nihil in intellectu nisi prius in sensu*'—there is nothing in the mind which was not first in the senses—says, exactly, that all our concepts are created, in one way or another, from elements provided by the senses. Far from being a Challenge, this is a Commonplace of Western Thought, and has been for a very long time. If embodiment says nothing more than this, then it looks much more like part of the problem than part of the solution.

So what else might 'embodiment' mean? Lakoff and Johnson say that concepts reside in minds, but they also claim to be physicalists, that is, to hold that 'there is a material basis for all entities taken as real within any scientific theory (p. 114).' This physicalism is possible, they say, because cognitive science operates at three levels: neural, conscious, and unconscious (see p. 102f). Here they seem to be drawing on the ideas of Donald Davidson, who has argued for a position called 'anomalous monism.' On Davidson's view, mental events such as onsets of belief must be identical on a case by case basis with brain events (hence 'monism'). But in spite of these case by case identities, mental events are not related in any law-like way to brain events (hence 'anomalous').⁶ It's a natural consequence of this view that generalisations which apply to beliefs and other cognitive entities may be impossible to capture at the level of brain events, not just because of practical difficulties, but in principle. A complete list of instances of the belief that it will rain tomorrow may have useful features in common, for example, reference to a future time. By contrast, a complete list of all the brain events which constitute the 'material basis' for these same beliefs, might have no useful features in common at all. Accordingly, if we want to retain the useful generalisations which involve cognitive entities, we have to accept these entities as irreducible: 'there are truths at each level that cannot be stated adequately at some other level (p. 110).' On this view, we cannot even in principle discard the entities of the conscious and unconscious levels in favour of a completely neural ontology.⁷

⁶ See Davidson (1969) and the ensuing literature.

⁷ In the same way, Chomsky hopes that mental representations of rules etc. will have a description at the level of atoms or molecules, and another at the level of neurons, 'without expecting operative, principles and structures of language and thought to be discernible at these levels Chomsky (2000: 25).'

Lakoff and Johnson believe that entities at all three levels—neural, conscious and unconscious—deserve to be called ‘real’ since they all have causal powers (p. 113–4). They are firmly committed to the reality of things like ‘verbs’, ‘concepts’, and unconscious semantic ‘frames’. But ... ‘not in the same sense as a commitment to the reality of chairs and rocks (p. 111).’ Chairs, rocks and neurons are the material basis: ‘Ultimately, the brain is all neurochemistry and neurophysiology (p. 113).’ We might call this a ‘really-*really*’ strategy. Concepts and frames really exist—but neurons *really* exist.

At this point, a natural question is: if concepts and frames have causal powers (as Lakoff and Johnson claim on pp. 116–7), are these causal powers *real*, or only real? None of the examples given on pp. 116–7 are of ‘causes’ from a cognitive level affecting the basic material level: frames causally generate inferences, image schemas causally bring about understandings, conceptual metaphors cause us to conceptualise things in certain ways. This very much suggests that higher-level entities do not causally influence the material base. In short, their causal powers are real (as they are themselves), but not *real*.

Worried, perhaps, that cognitive science may now look like some kind of ‘soft’ science, or that cognitive entities are beginning to look a bit irrelevant, Lakoff and Johnson try to argue that

explanation and motivation flow in both directions. To explain how the neurochemistry and neurophysiology function in networks of neurons, we need a theoretical level of neural computation. Explanation of what the physical neurons are doing flows from the middle level to the bottom [material] level (p. 113).

But this is simple equivocation. The ‘middle level’ of p. 113 is no longer the ‘phenomenological level’ of p. 103, populated by conscious mental states. Instead, it has been transformed into a level of ‘neurocomputational theory,’ populated by ‘neural gates, synaptic weights, thresholds and mathematical operations (p. 111).’ This neurocomputational theory organises hands-on research, and in that sense motivates and provides explanations for experimental work. But this is not at all to say that conscious mental states and unconscious hypothetical entities like image schemas and frames have causal influence on events at the bottom, material, level.

So there is a problem here, one with a familiar look: mental states and unconscious entities cannot *really* exist, in the way that chairs and rocks and neurons do, because they are not material. But if they do

not *really* exist then they cannot causally affect anything which does *really* exist. Concepts and frames, for example, cannot causally affect a physical process such as speech. Inferences and understandings cannot causally influence actions. In the end, the three levels story—even with its wobbly middle level—is just a complicated version of the identity theory. It is therefore useless as a way of explaining mind/body interaction. In addition, as Lakoff and Johnson use it, it implies that cognitive entities and processes do not after all have a *real* causal role.

So far, we have looked at two things which might be meant by 'embodiment'—good old-fashioned empiricism, and a Davidsonian identity theory. Neither of these helps. So what else might Lakoff and Johnson mean by saying that the mind is essentially embodied? Here's another line:

What we call 'mind' is really embodied. There is no true separation of mind and body. These are not two independent entities that somehow come together and couple. The word *mental* picks out those bodily capacities and performances that constitute our awareness and determine our creative and constructive responses to the situations we encounter (p. 266).

Now this certainly does help with the mind/body problem. If the word 'mental' simply picks out certain *bodily* capacities and performances, then mental causes are simply bodily ones. Unfortunately, this view already has a name. It's called behaviourism. It was supposed to have been swept away for ever by the cognitive revolution, which held that in order to *explain* our bodily capacities and performances, we need to invoke mental entities. For any cognitive scientist, the word 'mental' picks out something which *explains* the bodily capacities and performances which *reveal*—not 'constitute'—our awareness.

If this behaviourist excursus is a lapse, then the embodiment thesis must essentially be a form, possibly a developed form, of traditional empiricism. Any interest it possesses must therefore lie in finding the influence of sense-experience, not only in the construction of concepts but in the construction of metaphors, or to put it another way, not only in the content of concepts but in the (metaphorical) *connections* between them.

3. Conceptual Metaphor

Lakoff and Johnson's idea of '*conceptual metaphor*' seems intended to connect in some important way with this notion of conceptual structure, so we can continue our search for an original and useful meaning for the embodiment thesis by way of another of their Challenging claims, that abstract thought essentially involves conceptual metaphor. This claim is supposed to be particularly Challenging to analytical philosophy: 'analytic philosophy could not have sanctioned the existence of conceptual metaphors, and no future version ever will. It would mean giving up all of analytic philosophy's central ideas (p. 255).'

If this is true, then whatever conceptual metaphors are, they must be something more than ordinary, common-or-garden metaphors. Analytical philosophers do of course admit that ordinary metaphors exist, and there is a fair-sized debate in progress about how they work. It is true that the anglophone tradition has been wary of metaphor—but that is precisely because it has been so much aware of their power.

Let me give a particularly relevant example of the danger. Aristotle thinks of perception as analogous to the process by which wax receives an impression from a seal. The hot wax is able to take an impression of any shape because it has no intrinsic form of its own, and by the analogy, whatever it is in us which receives perceptions (since it can perceive any colour, sound, etc.) must have no intrinsic form of its own. Thus, says Aristotle, if knowing is like perceiving, then that which knows must also have no intrinsic form of its own. But if that which knows is 'mixed' with the body, it would have form, because all parts of the body have form. So it is reasonable to suppose that that which knows is not mixed with the body but distinct from it.⁸

Faced with an argument like this, whose persuasiveness depends entirely on the basic analogies (perceptions with wax impressions, knowing with perceiving), the natural reaction is to try to make the analogies as explicit as possible. The 'hostility' to metaphor one finds in Hobbes or Locke is not hostility to reasoning by analogy as such, but a careful preference for the more explicit simile.

Lakoff and Johnson's '*conceptual metaphor*' must therefore also be

⁸ See Aristotle (1993 reprint: Bk 2 ch 12 and Bk 3 ch 4).

something more than analogical inference, because analogical inference is of course well known to, and widely used by, analytical philosophers. Hume makes explicit the metaphor in the phrase 'matters of taste', and uses it to further his claim that there is an objective basis for aesthetic judgement. To take a more recent example, Searle's Chinese Room argument is obviously an argument by analogy: a computer processing a natural language is supposed to be *like* a human being sorting symbols in a room. In fact, Lakoff and Johnson clearly share the anglophone suspicion of inexplicit analogy, since they take Searle to task (pp. 261–5) precisely for ignoring some disanalogies they regard as important. 'Conceptual metaphors,' then, must be something more than ordinary metaphors, and something more than analogical reasoning, since analytical philosophers clearly acknowledge and employ both of these.

Let's begin by trying to say, in some very general, uncontroversial way, what we ordinarily take a metaphor to be. A metaphor, we usually suppose, is a figure of speech in which X is said or assumed to be identical with Y, though in fact, X merely resembles Y in some more or less complex way. The neck of a bottle is not a neck, but it's like a neck in a number of ways. Orsino's mind is not an opal, but Feste says it *is* an opal because he claims to see resemblances between it and an opal.⁹ Dying is not going on a journey, but numerous euphemisms say or imply that it is, hinting that we might meet the voyager again, when we travel the same road. Now, what makes conceptual metaphor so radically different from these common-or-garden instances?

Conceptual metaphors are supposed to be those which guide conceptualisation, 'as when we conceptualize understanding an idea in terms of grasping an object and failing to understand an idea as having it go right by us or over our heads (p. 45).' Lakoff and Johnson say, for example, 'The cognitive mechanism for such conceptualizations is conceptual metaphor (p. 45).' As we saw above, conceptual metaphor is alleged to be a *cause* of conceptualisations (p. 117).

It's natural to inquire here how the alleged effect is to be identified, independently of the alleged cause. How do we *know* that someone is

⁹ Shakespeare (1998 reprint: II.iv.74).

conceptualising, for example, understanding an idea as grasping an object? Mostly, we know because they actually use that kind of metaphor. They spontaneously say 'Now I've got it!' They respond appropriately to 'Hold on to that idea for a moment,' and so on. If we ask them explicitly whether they think understanding an idea is like grasping an object, they may very well say yes.

But if we don't have any *other* way of identifying occurrences of conceptualisation, then in operational terms, conceptualising A as B just *is* using the 'A is B' metaphor. And in that case, the supposedly empirical discovery that conceptual metaphors cause conceptualisations, amounts only to the tautology that when people use the metaphor, they use the metaphor.

There must, therefore, be other ways to establish that someone is conceptualising A as B. Suppose, for example, that a person makes a grasping gesture when claiming to understand an idea, or draws an 'over my head' picture when asked to depict failure to understand. Does this allow us to identify conceptualisation independently of the use of the metaphor? Hardly. As far as we know, these pictures and gestures are simply non-verbal uses of the metaphor.

The first problem, then, is that Lakoff and Johnson have to provide some operational means of identifying conceptualisations independently of uses of the corresponding metaphor, if the claim that the latter cause the former is to be taken seriously. Unfortunately, Lakoff and Johnson concede the impossibility of giving independent operational meaning to conceptualisation (on pp. 71-2), without seeming to realise what this does to their claim that 'conceptual metaphor ... plays a causal role in something real that we do, namely conceptualise (p. 117).' Not only does it reduce the causal claim to a tautology, it also means that there is no account, so far, of what a conceptual metaphor is (and so, still no account of what 'embodiment' means).

Notice, by the way, that the identification problem also arises in reverse. For Lakoff and Johnson, it is not the metaphor as an abstract entity but the individual's *cognition* of it which causes conceptualisations. But how are these cognitions of metaphors to be identified independently of the conceptualisations they are supposed to cause? What we have here is a supposedly causal, scientific, claim in which neither the cause nor the effect can be reliably identified.

A second obstacle to understanding what a conceptual metaphor is supposed to be, is that many of Lakoff and Johnson's examples of

metaphors are not metaphors at all. They say, for example, that for an infant, the subjective experience of affection is typically correlated with the sensory experience of warmth, the warmth of being held ... associations are automatically built up between the two domains ... these persisting associations are the mappings of conceptual metaphor (p. 46).

It's surely clear that these associations are not in any ordinary sense of the term metaphorical. Contrast another of Lakoff and Johnson's examples—the idea that life is a journey (p. 59). In this case, there is a series of resemblances, and the result is clearly a genuine metaphor, though presumably without any infantile association between the two domains.

Let me sum up the situation so far: when we looked at various things which 'embodiment' might mean, it seemed that our best hope of finding something interesting or new in 'embodiment' lay with the idea of conceptual metaphor. But conceptual metaphor itself has, so far, no clear meaning. It has not been distinguished from the conceptualisations it is supposed to cause, and it has no distinctive range of examples to provide even a preliminary understanding.

This brings us to Lakoff and Johnson's assertion that a metaphor is a 'mapping' between two 'domains,' consisting of associations between items in the two domains, and capable of further extension. Following this lead, we might try to understand *conceptual* metaphor as an unusually extended and extendable mapping between domains, noting that we use the mappings we do because we have the bodies we have, and being duly impressed by the depth and scope of the influence of these mappings on our thinking.

Unfortunately, so far from being new or Challenging, all this is evident from the earliest records we have. When the Psalmist says 'The Lord is my shepherd', he establishes a mapping between the Lord's relationship with His chosen people, and the shepherd's relationship with his flock. The metaphor is extended in various ways in the original, and the working out of further points continues to this day. The whole Song of Solomon, according to one (old) line of interpretation, is a working out of the 'God is a bridegroom' metaphor. The New Testament Parable of the Sower is interpreted explicitly, point for point, by Matthew, Mark and Luke. Does anyone doubt the far-reaching influence of these systematic or extended metaphors on Jewish or Christian thought? Does anyone suppose that creatures who had

no bodies, and who therefore neither farmed nor married, would have devised these metaphors?

In secular literature too, extended metaphor is an ancient and well-understood device. To take a comparatively recent example, Spenser, in a letter to Sir Walter Raleigh, calls his *Faerie Queene* 'a continued Allegory' and explains its moral and political intentions. Spenser clearly wanted his network of connected metaphors to influence his readers' ideas about gentlemanly behaviour, and about Elizabeth I. In the eighteenth century, Francis Hutcheson noted a kind of 'serious wit' in which 'resemblance is carried on through many more particulars than we could have at first expected.' If extended metaphor gives the core of the claim that mind is essentially embodied, then, since extended metaphor is one of the simplest and oldest instruments of criticism and hermeneutics, the embodiment thesis turns out (again) to be much more of a Commonplace than a Challenge.

Thus, if conceptual metaphor is to be original or interesting, its distinctive meaning must lie specifically in the terms 'mapping' and 'domain' themselves. Our search for a clear meaning for the idea of conceptual metaphor, and with it, an original meaning for the embodiment thesis, has brought us at last to these quasi-mathematical terms.

First, then, what do Lakoff and Johnson mean by a 'domain'? Lakoff and Johnson provide various examples of domains: 'Metaphor allows conventional mental imagery from sensorimotor domains to be used for domains of subjective experience. For example, we may form an image of something going by us or over our heads (sensorimotor experience) when we fail to understand (subjective experience) (p. 45).' Other examples of domains are: the subjective experience of affection, the sensory experience of warmth, the sensorimotor concept of verticality, the subjective judgment of quantity, knowing, seeing, importance, size, happiness, bodily orientation, similarity, proximity, abstract unifying relationships, experience of physical objects, thought, language ... the list goes on and on, and in the end it seems that a domain is just *anything* that can be supposed to have as members the items that figure in a mapping. In this case, the idea of a domain does nothing to help explain what a metaphor or a mapping is. On the contrary, we are supposed to be able to pick out cases of mapping first. What, then, is a 'mapping'? Unfortunately, a mapping, in turn, seems to be any sort of correlation of anything at all. Open a dictionary at random and pair the top word of the left hand page with the top word of the right

hand page. Now pair the second top word of the left hand page with the second top word of the right hand page. Keep going until you realise the futility of it all. The result is not in the least metaphorical—but it certainly *is* an extended and extendable mapping. There is nothing in Lakoff and Johnson's use of the terms 'mapping' or 'domain' to bring out what is distinctive even about ordinary metaphor, and *a fortiori*, nothing which explains the radically new and Challenging conceptual metaphors they claim to perceive.

The result is that the bold claim that abstract thought is essentially metaphorical (the second of Lakoff and Johnson's Challenging theses) has simply vanished into smoke. It's also worth noticing that the terms 'domain' and 'mapping,' if they are to serve in an empirical science, have to be given some *empirical* meaning: without some criteria for identifying mappings in real life, anyone can start seeing mappings wherever he or she likes (rather as they did with 'drives' in the first half of the twentieth century). However, even if endowed with empirical meaning, the terms would still not tell us what is special about conceptual metaphor, since perfectly ordinary metaphors are also supposed to involve 'mappings' between 'domains.'¹⁰

So far, we have looked at two of the three claims which, according to Lakoff and Johnson, are going to revolutionise Western Thought. The claim that the mind is essentially embodied reduced either to behaviourism, or the identity theory, or simple empiricism, with only the idea of a radically different kind of metaphor called conceptual metaphor to maintain its pretensions to originality. The second claim, that most abstract thought is metaphorical, also depends entirely on the idea of conceptual metaphor. But conceptual metaphor reduces to the very familiar notion of extended metaphor, a collapse disguised only by

¹⁰ Fauconnier (1997) maintains a similar reticence about what mappings are, telling us only that 'A mapping, in the most general mathematical sense, is a correspondence between two sets that assigns to each element in the first a counterpart in the second.' This is a bad definition for the mathematical case, since it implies that mappings must be exhaustive and one-to-one, and it's a bad definition for Fauconnier's own purposes, since he wants to allow mappings within a single domain (p. 102f). But worst of all, it's empirically empty. The word 'mapping' is like the word 'thing': to say that something happens in a certain way because of a mapping is no more explanatory than saying that something acts in a certain way because it's a thing.

Lakoff and Johnson's consistent vagueness about the key terms—conceptualisation, metaphor, domain, mapping. The third claim, that most thought is unconscious, is increasingly suspect but not original to Lakoff and Johnson, and I shall say no more about it here.¹¹

The result is that we find ourselves thrown back on the catch-22 problem of inescapable but conflicting metaphors. If it is true, as Lakoff and Johnson claim, that philosophical theories simply choose among the available metaphors, which then 'have the *causal effect* of constraining how you can reason (p. 117),' it follows that we are *less* well placed to make any sort of rational choice between theories. The greater the stress on the metaphorical nature of language, the more intractable the mind/body problem becomes.

Lakoff and Johnson often refer to what they take to be empirical results, and though they sometimes regard these as metaphorical by nature too (since even experimental results about axons and dendrites, for example, will inevitably be expressed in terms of the 'Neural Computation metaphor,' p. 111), their general tendency is to regard empirical results as a means of deciding to favour one metaphor, or some of its 'entailments,' over others. In other words, their general tendency is to suppose that we can get access to stable facts, which ought to guide our inevitably metaphorical talk about the mind. This means that the problem is to be solved for us, simply, by new and better facts. Whether this optimistic view is true or false, it cannot be said to be original, or to depend in any way on the claim that abstract thought is mostly metaphorical.

¹¹ Lakoff and Johnson (1980) characterised metaphor as a method of understanding A in terms B. Does this provide a clue as to what a metaphor is? The short answer is no. We understand temperature in terms of the average kinetic energy of molecules. Is average kinetic energy a metaphor for temperature? We understand the Gulf War in terms of the supply and demand of oil, but the latter is not a metaphor for the former. Nor does every metaphor improve or significantly alter our understanding. Tables and chairs have supports to raise the main surface to a convenient height above the floor. Whose understanding of these supports is improved by calling them *legs*?

The general usefulness of metaphor in developing new ideas has been recognised at least since Aristotle. But this is only the beginning of the question. What we want to know is *how* metaphor does this. Unfortunately, neither 'conceptualisation,' nor 'mappings' and 'domains,' tells us anything about what metaphor is or how it works.

4. Metaphor in Philosophy

It seems, then, that the embodiment claim and the claim that thought is metaphorical tell us nothing useful or new. This is a disappointment, but it is not the last disappointment this large book has to offer. Lakoff and Johnson also set out to document examples of extended metaphor in philosophy. As a method of exegesis, this is hardly likely to offer a Challenge to Western Thought, but it is not impossible that it could throw genuine light, here and there, on particular philosophical arguments.

Since our main concern has been the mind/body problem, we can turn for an example of the method in action, to Lakoff and Johnson's remarks on metaphor in Descartes. Our last hope of help with the mind/body problem, from Lakoff and Johnson, is their analysis of Descartes' extended metaphors. They claim that

What makes [Descartes'] argument fit together, make sense, and seem intuitively appealing ... is a peculiar metaphorical logic in which Descartes has pieced together a number of common metaphors about the mind and then followed out their entailments ... without the metaphors, the logic of the argument is simply not present; nor is the metaphysics of his theory of mind (pp. 392–3).

They notice that Descartes adopts a 'Knowing As Seeing' metaphor. This metaphor goes back at least to Plato, and it plays an important role in the famous Allegory of the Cave, which is set out (and explained explicitly as an extended metaphor, point for point) at the beginning of Book VII of *The Republic*. According to Lakoff and Johnson, the visual metaphor is 'so central to our conception of knowledge that we are seldom aware of the way it works powerfully to structure our sense of what it is to know something (p. 394).' This being so, it's easy for Descartes to assume that 'metaphorical objects (our ideas) are illuminated by an inner light (the 'Natural Light of Reason') and are observed by a metaphorical spectator (our faculty of understanding) (p. 393).' 'Reason ... is conceptualized by the Knowing Is Seeing metaphor as a person who can see (p. 395).'

Lakoff and Johnson claim that 'Descartes' reasoning does not go through without what we know about illumination and light from the source domain of vision (p. 396).' For example, his conclusion that the mind can know its own ideas with absolute certainty only follows

because we take it that a person with normal vision looking at an object in good light can see it with absolute clarity. They state flatly that 'Descartes' argument for this conclusion cannot be made without the metaphors just discussed (p. 396).' Another of Descartes' conclusions—that all thought is conscious—arises from the same metaphors, though it 'has been invalidated by virtually all of cognitive science (p. 396).'

Let's consider these two 'conclusions' of Descartes' in turn. To say that the mind can know its own ideas is ambiguous, between knowing what they are and knowing that they are true. If it means knowing what they are, Descartes in fact stresses how difficult it is for the mind to know its own ideas. Many of our ideas are confused and unclear, and the most strenuous corrective efforts may be needed to know them adequately.¹²

Contrary to Lakoff and Johnson's view, this lack of perfect clarity might well be an 'entailment' of the visual metaphor, since vision is often impaired in various ways, and some of the things we see, such as clouds, just *are* diffuse. But in fact it's obvious enough without the visual metaphor.

People often misunderstand ideas, often reason badly, often explain what they mean inaccurately, and so on. In these cases, Descartes is very much aware, we do not know our own ideas with anything like the clarity we want. It is *because* this fact is obvious enough without the visual metaphor that the visual metaphor appeals.

(There are also ideas which any normal person can competently explain, reason with, and understand. In our dealings with ideas, some strike us as difficult, others as comparatively easy. This is not something irreducibly metaphorical, but the plain fact of our variable competence in reasoning, explaining, and using language. The visual metaphor appeals *because* of the facts, rather than somehow constitut-

¹² Descartes says, for example, "I term that 'clear' which is present and apparent to an attentive mind, in the same way that we see objects clearly when, being present to the regarding eye, they operate upon it with sufficient strength," *Principles of Philosophy* I, 45–6.

In this definition of clarity, Descartes obviously draws on the visual analogy, and consistently with it, implies that some ideas are not clear. Knowing our own ideas adequately and clearly is as difficult as it often is to get a clear view of something.

ing them.)

Turning to the second possible meaning, that the mind can know the truth of its own ideas with absolute certainty, the fact is that Descartes' argument, far from being irreducibly metaphorical, is a straightforward argument by example. There is an *example* of a truth, according to Descartes, which can be known with absolute certainty—the truth 'I am thinking.' When anyone formulates this thought, he or she can be absolutely certain it is true. So in this case at least, Lakoff and Johnson misrepresent Descartes' conclusions, and exaggerate the role of the 'thinking is seeing' metaphor in producing them. Sadly, the same problems beset their claim that for Descartes, all thought is conscious.

Like any believer in innate ideas, Descartes is inevitably committed to unconscious ideas, since a child or uneducated adult is supposed to possess all kinds of ideas which he or she has no conscious inkling of. This was already clear from Socrates' famous conversation with the slave boy, who is supposed to have 'latent' geometrical knowledge all along, which gradually becomes conscious as a result of Socrates' questions. Descartes says, for example,

The idea of God is so imprinted on the human mind that there is no one who does not have in himself the faculty of knowing him; but this does not prevent many people from being able to pass their whole lives without ever distinctly representing this idea to themselves.¹³

It's true that Descartes defines thought (in the *Principles* and in the *Second Replies*) as whatever we are immediately conscious of, and there are, as a result, serious exegetical problems about how he sees the relationships between thought and consciousness, thought and thoughts, thoughts and ideas. But the above passage clearly shows that Descartes holds—as he must—that some of our innate ideas may well remain 'latent.'

Lakoff and Johnson, however, think Descartes had to believe that all thought is conscious because his metaphors forced him to. They say, 'If all ideas are objects and all objects can be seen consciously, and if Knowing Is Seeing, then all ideas can be known consciously. The no-

¹³ From a letter to Clerselier, 17th Feb, 1645. See Adam and Tannery (1976: vol. 4 p. 187)

tion that there could be an idea or thought process that could not be accessible to consciousness would be like an object that was by nature invisible (p. 397).’ But even if this metaphor had Descartes in its vice-like grip (inconsistently with Lakoff and Johnson’s claims elsewhere that we can choose between metaphors), what is to prevent him from conceiving of a realm, like the bottom of the sea, or the far side of the moon, which he cannot see? The visual metaphor obviously does not compel us to think that everything is visible to us—quite the reverse.

Even if the demand is for an object which is *by nature* invisible (not that this is necessary for a concept of the unconscious), the fact is that Descartes clearly *believes* in such objects. Angels and spirits are essentially invisible, for Descartes, and so are the smallest material particles. Lakoff and Johnson’s claim, that the visual metaphor compels Descartes to hold that all thought is conscious, is false in just about every way it could be—false to Descartes, false to the metaphor, and false to the metaphor’s influence on Descartes.

There are of course cases in which focussing on the metaphorical content of a philosophical view might be illuminating, if one is willing to read carefully and think hard. Aristotle’s wax impression is a case in point, and Hume’s ‘matters of taste’ is another. Lakoff and Johnson, however, seem to be dazzled by their great idea, the idea of conceptual metaphor. This is a pity, because to the extent that it is meaningful, the idea is banal.

5. Conclusion

We may conclude that:

- the three-level version of the identity theory is no better than any other at explaining how a non-material entity or process could have a causal influence on material entities,
- conceptualisation is not distinguishable from metaphor use (and so not caused by it),
- embodiment and conceptual metaphor are simply new names—worse than the old names because less clear—for empiricism and extended metaphor,
- no useful idea of metaphor is to be found anywhere in Lakoff and Johnson’s writings, because the key terms (‘domain,’ ‘mapping’ etc.) are so vague as to include any sort of correlation whatsoever, and

- the claim that philosophy has metaphorical content is neither surprising nor (as Lakoff and Johnson employ it) illuminating.

Let's return, then, to the problem with which we began, the problem of understanding how a mental entity or process could be a cause. Perhaps it's time to contemplate the awful possibility that cognitive explanation might not be causal, and ask ourselves what might be achieved with a more modest, non-causal, cognitive science.

One very natural idea is that the cognitive scientist might hope to organise the data in ways which would fruitfully guide neural research.¹⁴ It might be, for example, that the (very heterogeneous) raw oscilloscope readings of the sounds someone produces and recognises as the definite article, can be unified by supposing that he or she has a Saussurean 'sound-image' in mind. If some such unconscious concept resides in the mind, and if the mind has a complete material basis in the brain, then it is natural to expect to find *some* sort of recognisable correlative of the sound-image in the brain. One might therefore advise any neural researcher at a loss for something to do, to go and look for the neural realisation of the sound-image. It must be in there somewhere.

If this works, it works, and is to be encouraged. My own guess is that there will be many cases—like the memory 'engram'—where these hypothetical unconscious entities will prove to be illusory. Entirely plausible, as the engram was, and the more or less accidental stimulus for other discoveries, but basically a waste of time. Seeming 'hits,' when they occur, will tend to owe as much to changes in our understanding of the cognitive entity we're looking for, as to straightforward neural discoveries.

These points derive from the idea that cognitive entities *really* are not causes. If they really aren't, they will if anything follow rather than lead causal investigation. In short, the hope that cognitive explanations might fruitfully guide neural research is a way of disguising the causal ambition of cognitive explanation, rather than abandoning it.

So if we really do abandon a causal role for cognitive entities and

¹⁴ Chomsky (2000: 25–26), for example, says, 'Perhaps C-R [computational-representational] theories will provide guidelines for the search for such mechanisms [of the brain].'

processes, what usefulness might they retain? Wittgenstein says (PI 156), 'these mechanisms are only hypotheses, models designed to explain, to sum up, what you observe,' and this points towards a kind of explanation which sums up, or finds useful patterns in, the raw data.

Consider this example. Music shops organise their CD's by style of music (Rock, Classical, Jazz, Folk etc.) and within styles by the performer's or composer's surname. These methods of organisation, though so useful as to be virtually universal, ignore causal relationships. No one would argue that all CD's of music by Haydn, Hindemith and Hildegard of Bingen must have some distinctive physical feature in common. A Martian scientist, investigating the physics of CD's, would not be well-advised to look for something on the pitted surface of the disc which makes it into a Jazz or Baroque CD. In spite of this, these methods of classification are not only useful in practical terms, they also provide a basis for generalisations which we find interesting, such as that Jazz uses more syncopation than Rock, or that Baroque music fundamentally expresses balance.

The same kinds of classification can be seen in every library, and they show at least, that non-causal classification is perfectly familiar, that it can be illuminating given our interests, and that it can deserve its place even if it is useless—in fact even if it is positively misleading—to people researching the physical basis.

On the other hand, there will be cognitive linguists who feel that, if classifying books or CD's is not like classifying stars or butterflies (because not preparatory to physical research), it is unscientific, and must therefore represent a demotion. Frankly, utterances are much more *like* books or CD's than stars or butterflies, because the important thing about them is that they have meaning. Anyone who feels that this makes the study of meaning unscientific, because not causal, might find it helpful to ask how far the discovery of causes could go without a meaningful language for comparing and communicating results.

Whatever we do about the demotion feeling, however, we should at least say clearly whether cognitive explanations aspire to causal status—and if they do, we should explain how something non-material can be a cause. The days of the identity theory fudge are numbered.

REFERENCES

- Aristotle *De Anima*, ed. by Michael Durrant, Routledge, London and New York, 1993.
- Chomsky, Noam (2000) *New Horizons in the Study of Language and Mind*, Cambridge University Press, Cambridge.
- Churchland, Patricia (1989) *Neurophilosophy*, MIT Press, Cambridge, MA.
- Davidson, Donald (1969) "Mental Events," reprinted in *Actions and Events*, Oxford University Press, Oxford, 1980.
- Descartes, Rene *The Philosophical Writings of Descartes*, ed. by J. Cottingham, R. Stoothoff and D. Murdoch, Cambridge University Press, Cambridge, 1985, *Oeuvres de Descartes*, ed. by Adam and Tannery, Librairie Philosophique J Vrin, 1976.
- Fauconnier, Gilles (1997) *Mappings in Thought and Language*, Cambridge University Press, Cambridge.
- Hobbes, Thomas (1651) *Leviathan*, Penguin, London, 1968 (see for example Chapters 4&5).
- Hume, David (1741) "Of the Standard of Taste," *Essays, Moral, Political and Literary*, ed. by E. F. Miller, Indianapolis, 1987.
- Hutcheson, Francis (1725) "To the Author of the Dublin Journal," *Dublin Journal*, see *Philosophical Writings* ed. by R. S. Downie, London, 1994.
- Jackendoff, Ray (1990) *Consciousness and the Computational Mind*, MIT Press, Cambridge MA.
- Lakoff, George and Mark Johnson (1980) *Metaphors We Live By*, University of Chicago Press, Chicago.
- Langacker, Ronald (1990) *Concept, Image and Symbol: The Cognitive Basis of Grammar*, Mouton de Gruyter, Berlin.
- Locke, John (1689) *An Essay Concerning Human Understanding*, ed. by A. C. Fraser, Dover, 1959 (see esp. Bk 2 Chapter 11, and Bk 4 Chapter 17).
- Pinker, Steven (1994) *The Language Instinct*, William Morrow and Company.
- Plato, *The Republic* Book 7, trans. by B. Jowett, Sphere, 1970.
- Reid, Thomas (1764) *An Inquiry into the Human Mind on the Principles of Common Sense*, ed. by R. E. Beanblossom and K. Lehrer, Hackett, 1983.
- Reid, Thomas (1785) *Essays on the Intellectual Powers of Man*, ed. by R. E. Beanblossom and K. Lehrer, Hackett, 1983.
- Searle, John (1980) "Minds, Brains and Programs," *Behavior and Brain Sciences* 3.
- Shakespeare, William *Twelfth Night* in *The Complete Works*, ed. by S. Wells and G. Taylor, Oxford University Press, Oxford, 1988.
- Sweetser, Eve (1991) *From Etymology to Pragmatics*, Cambridge University Press, Cambridge.
- Wilson, Brendan (1998) *Wittgenstein's Philosophical Investigations: A Guide*, Edinburgh University Press, Edinburgh.
- Wittgenstein, Ludwig (1933) *The Blue and Brown Books*, Blackwell, Oxford,

1958.

Wittgenstein, Ludwig (1953) *Philosophical Investigations*, Blackwell, Oxford, 1991.

brendan@boz.c.u-tokyo.ac.jp