Part II

The Embodied Mind, and How to Live with One

George Lakoff
3
How the Body Shapes Thought: Thinking with an All-Too-Human Brain

George Lakoff

Why It Matters

The mind isn’t what we thought it was. Philosophy wasn’t even close in its speculations.

For over two thousand years, philosophers have mostly viewed the mind as disembodied. The disembodied mind was not an empirical discovery, but rather a philosophical creation—an a priori philosophical doctrine that has shaped our understanding of ourselves and our world. Even the early development of cognitive science from formal logic, artificial intelligence, and generative linguistics was itself shaped by those a priori philosophical views. First-generation cognitive science was more philosophical than empirical.

But over the past two decades, something truly remarkable has happened. With the help of neuroscience and a truly empirical cognitive linguistics, cognitive science has transcended its own a priori philosophical beginnings. A new understanding of the mind has emerged, empirically based and freed from age-old philosophical baggage. The new view of mind changes everything—massively, in an almost shocking way.
The magic lamp of science has been rubbed and the genie has exchanged the old view of mind inherited from the philosophers for a new one: in the new view, thought is largely unconscious, embodied, and metaphorical.

Where in the old view, mind was symbolic, modelled by the algorithmic manipulation of abstract symbols, as in a digital computer or a logical deduction, thought in the new view is biological and neural, not a matter of symbols.

Where, in the old view, thought was disembodied and abstract, thought in the new view is embodied – physical in nature, with concepts precisely and exquisitely sculpted by neural circuitry that evolved to run a body: neural circuitry for vision, for bodily movement, for emotion, for empathic connection, and for functioning with a clunky physical body in a physical environment rife with peculiarities. The peculiar structures of our concepts reflect the peculiarities of our bodies, our brains, our interpersonal life, and our experiences.

Where mind was seen only as conscious in the old view, in the new view it is mostly – my guess is 95 per cent or more – unconscious, that is, not directly accessible to conscious introspection. We do not, and cannot, directly know our own minds. Descartes was in error. Phenomenology, which depends on conscious introspection, has real utility as a guide to mind, but it is limited to the conscious tip of the iceberg of the cognitive unconscious.

Where the old view saw thought as literal and consistent – a proper topic for logic – in the new view, thought is literal only in part. So-called ‘abstract thought’ is largely metaphorical, making use of the same sensory-motor system that runs the body.

The genie is out of the bottle and it’s not going back in. The change from a mind that is disembodied, literal, and conscious to one that is unconscious, metaphorical, and shaped by the body is staggering and all-pervasive. Once you recognize the truth and import of these discoveries, you can’t go back to believing what you used to believe.

The central philosophical ideas – Time, Events, Causation, Morality, the Self, and even Being, Truth, and Knowledge, are not what they were thought to be, namely, literal and univocal. Each is a mostly metaphorical concept, drawing on the body for its logic; and each is multivocal, defined by many, mutually inconsistent metaphors.

Causation, for example, is not just a simple, literal, univocal aspect of nature. Causation is a human concept, not an objective feature of the world. There is not one causation, but nearly two
dozen, each with its own logic – and each metaphorical in a very different way. There are, of course, phenomena in the world that we can reasonably call ‘causal’. But they are not all of a piece; they work by different logics; we reason about them via different metaphor systems; and it is our minds, not the external world, that make them all forms of one thing: ‘causation’.

Central philosophical distinctions, like metaphysics versus epistemology, go out the window – as do other classical dichotomies: objective versus subjective, realism versus idealism, rationalism versus empiricism, Anglo-American analytic philosophy versus Continental philosophy, structuralism versus post-structuralism. In each dichotomy, both poles dissolve before the new conception of mind; each dichotomy was a mistake. The result: philosophy becomes more interesting, more challenging, and much more worthy of public attention. A new philosophy is urgently needed to keep up with the science. We have called such a philosophy ‘embodied realism’.

Was Wittgenstein right that philosophy is just figuring out language games and dissolving puzzles? Or is the correspondence theory of truth right, with the consequence that there are real philosophical problems because the world has an objective structure? The answer is, ‘Neither!’ From the perspective of the embodied mind and embodied realism, both views make empirically incorrect assumptions. Yes, embodied realism has real philosophical insights, dealing with real philosophical problems. But the insights are new and the problems are not the old ones. Philosophy benefits, because it is taken out of old, pointless entanglements that were based on an inadequate view of mind.

Even mathematics and the philosophy of mathematics benefit from the spread of cognitive science and neuroscience throughout our intellectual life. The old romantic mythology of mathematics went like this:

Mathematics is abstract, disembodied, and transcends the human mind, yet it is real. It has an objective, literal existence independent of any beings with minds. Indeed, mathematics structures the universe, and characterizes the very nature of rationality via mathematical logic, as well as the very nature of intelligent thought, via artificial intelligence.

It was a nice myth while it lasted, but it’s not true. When we understand mathematical ideas via cognitive science, every part of the old romance goes up in smoke. Poof! Mathematics is embodied – even the idea of infinity as we shall see. Mathematics does not
transcend real, embodied minds, and it is not purely literal – it is metaphorical, grounded in embodied experience. What does mathematics become? Not the measure of the universe, or of rationality, but something even more interesting, more beautiful, more challenging – and much more accessible and understandable.

A similar story holds for morality. Moral concepts too are embodied and metaphorical, and to understand this in full detail is to give up forever on the idea that there is a transcendent morality based on transcendent universal reason. After all, no part of our conceptual system is transcendent and there is no transcendent universal reason. There is more than one moral system that people use, and there are about two dozen fundamental metaphors for characterizing morality, metaphors that are commonplace around the world, but not consistent with each other.

For example, morality can be seen as a matter of empathy, care, and responsibility. Or it can be see as obedience to a moral authority.

Morality can be seen as a matter of accounting, of balancing the moral books – in one of a number of ways. Suppose someone harms you. You may find it moral to balance the books by seeking restitution, having him do something good to make up for the harm. Or one can find it moral to insist on retribution (harming him back), or revenge (taking something good away from him). Or one may find it moral to forgive the moral ‘debt’ – to wipe it from the account books.

For some, morality is toeing the line, staying on a moral path defined by rules and not deviating from them. For others, morality is a matter of balance and fairness, of finding a middle way. Still others find morality in community, in maintaining community norms.

It is common for some to find morality in the avoidance of immorality, with immorality seen as akin to a disease, a contagion that can be contracted through contact with immoral people. The moral solution is quarantining immoral people – imprisoning them, ghettoizing them, or stigmatizing them. Or one may find morality in purity, in avoiding all taints of the immoral and in regularly purifying oneself.

Others see evil as a tangible force in the world that cannot be avoided, and find morality in discipline, in being morally strong so as to stand up to evil.

For many people, certain of these metaphors define what morality is. And they define morality in different, often inconsistent, ways.
Does that mean that we give up on morality as a guide for our lives? Not at all! The cognitive science of moral ideas requires us, as we shall see, to become more morally sensitive, more insightful about the nature of morality, and more morally responsible. Understanding the embodied metaphorical nature of moral concepts makes moral understanding that much more urgent and far more possible.

The same can—and will in the next lecture—be said of both science and religion. The central theoretical concepts of science are not literal, disembodied, and objective. They too are embodied and largely metaphorical. Naive scientific realism cannot work. But embodied scientific realism can, as we shall discuss. Scientific laws are not in the world, nor do they objectively 'fit' the world. But that does not make science one bit the less valid when we understand embodied realism. To come to grips with the embodied metaphorical nature of scientific concepts is to give a new dimension to scientific theorizing, to understand science better and to understand the world more realistically.

Finally, religion. Every conceptualization of God is metaphorical. There are three basic classes of metaphors for God and they are very different from one another. When one understands just how all the various conceptions of God are inescapably metaphorical, then the question 'Does God exist?' becomes a strange question: God under which metaphor? To recognize that the question is inherently metaphorical is to know that no answer can be literal. That is what makes the question strange.

Is there a 'soul' that lives on after death? Obviously, cognitive science can't say for sure, but it can say more for sure than any a priori philosophy has. Cognitive science and neuroscience can tell us what properties are embodied and therefore what properties a disembodied soul cannot have. When we are done with our list in Chapter 4 of what a soul cannot be, our understanding of a 'soul' will be utterly and irrevocably changed.

Is spiritual life dismissed by such understandings? No. But it is a very different conception of spiritual life than one finds in anything like the most common versions of Western religions. Spiritual life is physical, utterly of the body, of this world not any other. This is an understanding of spiritual life that is inconsistent with most traditional interpretations of Christianity, Islam, and so on, yet it is focused as much—or more—on spiritual and moral practice.

That is where these lectures are going, towards embodied realism as a scientific, philosophical, moral, and spiritual undertaking all of a piece, undifferentiated, characterizing a way of life.
The Story in Brief: How the Body Shapes Concepts

The science behind these lectures is well known and well documented, backed up by an extensive and profound literature in cognitive linguistics, neuroscience, cognitive anthropology, and cognitive and developmental psychology. I have done my best to survey that literature, make it accessible, and contribute to it with the help of distinguished colleagues in such works as *Philosophy in the Flesh* (with Mark Johnson; 1999), *Metaphors We Live By* (with Mark Johnson; 1980), *Where Mathematics Comes From* (with Rafael Núñez; 2000), *More Than Cool Reason* (with Mark Turner; 1989), *From Molecules to Metaphor: The Neural Theory of Language* (with Jerome Feldman; forthcoming) as well as my own books, *Women, Fire, and Dangerous Things* (1987) and *Moral Politics* (1996, 2000). For the sake of these short chapters, I will review some relevant details as briefly as I can.

Brevity is best, because the details are somewhat dull in themselves, one by one. It is only when the details are put together in a larger picture that their truly radical consequences appear.

A World Without Colour

I used to believe in the correspondence theory of truth and other varieties of Anglo-American analytic philosophy. I was raised to believe in it. It was what my teachers told me back in the early 1960s and what all the latest books said back then. According to this theory, a sentence like ‘The chair is red’ is supposed to be true just in case the entity designated by ‘the chair’ is in the set of red things in the world.

The idea behind this theory of truth was, in its basic outlines, as follows. We start with a metaphysics: the world, at any time, is made up of objects, with properties, and with relations holding between them. Sentences of a natural language have ‘referring expressions’ that can designate objects, and other expressions that can designate properties and relations. Hence the expressions of a language can ‘fit’ the world as inherently structured.

In formal logic, the world as seen in this metaphysics is replaced by set-theoretical models. The entities in the world are modelled by the ‘abstract objects’ in the model; the properties of the entities in the world are modelled by sets of abstract objects in
the model; and the relations between entities in the world are modelled by sets of \( n \)-tuples in the model. This allows one to characterize a world-state. Intensional logics then go on to talk about relations between world-states, either complete or partial.

In my own early work on generative semantics, I assumed all this. I further assumed an early cognitive theory in the early 1970s, namely, that all mental models were partial world-state models. The idea was that language, via generative semantics, could fit the mental models and that the mental models could fit the world. It was an attempt at a correspondence theory of truth with the intervening step of partial mental models taken from logic.

When I found out that this whole picture was false, I almost fell off my chair. It was the summer of 1975. I was listening to a lecture by Paul Kay on how the neurophysiology of colour vision explained the Berlin & Kay (1969) results about the meanings of colour terms. The result was this: Each language has basic colour terms – single morphemes that exclusively designate colour, rather than say blood or grass. The number of such terms differs from language to language, from as few as two to as many as eleven. The portion of the colour field covered by such terms may differ considerably from language to language. *But the best examples of the terms are always the same focal colours, for example, focal red or green.*

This fact was explained by research on the neurophysiology of colour vision at Berkeley by De Valois and his colleagues (De Valois & Jacobs 1968; De Valois & De Valois 1975) Colour is a function of the following factors:

1. the reflective properties of objects;
2. lighting in the given area;
3. three kinds of colour cones in the retina, responding to high, medium, and low frequencies;
4. neural circuitry connected to the cones, with neurons in the brain responding maximally to certain wavelengths as filtered by the cones.

Though there are only three kinds of colour cones, the complex neural circuitry connecting them gives rise to *four* primary colours: red, green, blue, and yellow.

Focal colours (e.g., red or blue) correspond to neural activation peaks over the circuits. Basic colours that are not primary (e.g., orange) are produced when two or more circuits are simultaneously active and when each is maximally firing given the firing of the other (or others).
This explains the universality of focal colours, regardless of the number of basic colour terms. It also explains much of the internal structure of colour categories – what counts as 'light red' or 'deep red' or 'pink'.

What we learn from all this is that colour is not out in the world. Colour arises from the interaction of two factors in the world external to us – reflected wavelengths of light and nearby lighting – and two bodily factors, namely, colour cones and the neural circuitry connected to them. Rather, colour is embodied and interactional – what Locke called 'secondary'.

One cannot merely identify colour with wavelengths. Any colour can be produced by combining three different wavelengths of light – and this can be done with different combinations of three wavelengths. What the three wavelengths can do is to activate the colour cones to different degrees. The complex neural circuitry then produces an experience of a colour.

When I learned this, it became clear that the correspondence theory of truth is false. The sentence 'The chair is red' could not be true just in case the entity designated by 'the chair' is in the set of red things in the world – because there is no set of red things in the world independent of our bodies and brains and perceptual systems. To me this was a shock. What I had been taught and what was generally believed in the field of semantics was not true.

Other philosophical positions were disconfirmed by this fact as well. Plato's idealist view of abstract essences cannot work for colour, since colour is embodied, tied to colour cones and neural circuitry. The embodiment of colour also ruled out certain common Continental views, for example, that meaning is subjective or arbitrary, a matter of how disembodied signs were fitted to the world. The commonalities of normal human bodies – the same colour cones and circuitry – ruled out the view of meaning as purely subjective. Colour is neither purely objective (out there independent of us) nor purely subjective, in the sense of being just a matter of individual experience.

Next, we learn something deep about subjective experience. Our understanding of how colour is tied to neural activation tells us nothing about the subjective experience of colour, the colour qualia. Knowing which circuitry produces the experience of red does not explain – or even describe – the experience of red in itself. The quality of subjective experience cannot be explained by our theories of neural activation alone. On the other hand, the quality of our subjective experience of colour is tied to neural activation – colour doesn’t occur without neural activation and it changes
systematically as the neural activation changes. In short, the quality of our subjective experience is not disembodied, not floating in air. The quality of experience is tied to neural functioning. Descartes was wrong. There is no mind–body separation.

Moreover, the meaning of colour terms is not a matter of free-floating symbols to be fitted to an external reality. Fodor is wrong as well. Functionalism – Putnam’s early idea that one can study everything about the mind including concepts via abstract computation using abstract symbols – is also falsified. The meaning of colour terms is not in abstract symbols manipulated by algorithm. It requires embodiment – in the very neural circuitry linked to the colour cones. The first generation of cognitive science itself, which arose from a priori philosophy, was disapproved by neuroscience. Indeed, the science suggested a new philosophy, as we shall see.

Most philosophers consider colour an exceptional case. It isn’t. It is the typical case. Any concept at all must be neurally embodied. The reason is simple: we think with our physical brains. For any given concept, the question is not whether it is neurally embodied, but how.

**Basic-Level Categories**

Colour was just the beginning of my disillusionment with the philosophy I had been taught. In that same summer of 1975, I heard Eleanor Rosch give one of her first lectures on basic-level categories, which are mid-level categories such as chair, as opposed to the superordinate furniture or subordinate rocking chair – or car as opposed to vehicle or sports car. Rosch had demonstrated that basic-level categories have different cognitive properties than superordinate categories. They are defined by our capacities for gestalt perception, motor movement, and mental imagery. Compare chair and furniture. You can get a mental image of a chair, but not of a general piece of furniture (as opposed to a chair, bed, table or couch). You have motor schemas for interacting with chairs, but none for interacting with general pieces of furniture. In short, the basic level is the highest level at which mental imagery, motor schemas, and gestalt perception characterize the entire category and the basic level is the optimal level at which people interact with objects.

Brent Berlin has shown that, in a jungle culture where people can name over 800 plants, they are almost entirely accurate with
respect to biological taxonomies at the basic level (around the 90 to 95 per cent level), but their accuracy falls by half at the next lowest level. The basic level appears to be the level at which we have evolved to function optimally in the physical environment given our sensory-motor systems. It is also the level at which we learn earlier, remember best, and know most.

The basic level, which is a reality about human conceptual systems, is not a mere reflection of external reality, but a matter of our interaction with our environments. Like colour, it is embodied and interactional. Our concept of a chair has to do, after all, with our ability to sit, which has everything to do with our bodies. It is a fundamentally embodied concept.

Image Schemas

As concepts, spatial relations present vexing problems for any serious student of the mind. Back when I was still trying to make formal logic work for natural language semantics, I was stumped by spatial relations. In English, we have concepts corresponding to prepositions like ‘in’, ‘on’, ‘through’, ‘around’, ‘behind’, ‘in front of’, ‘above’ and so on. Each language has a different set, with a different range of meanings. Yet in each language, such concepts have a perceptual dimension. *They link what we see with what we know.* Given a scene with one object located with respect to another, we know which spatial relations term to use, whether more than one would work, how well they work if they don’t fit perfectly and what regions of space they cover. At the same time, we *reason* with spatial relations concepts, for example, we know that:

If A is in B and B is in C, then A is in C.

Moreover, we can use spatial relations to form mental images, and we can reason in terms of those images.

The dual perceptual and conceptual nature of spatial relations concepts raises the following questions:

How do you link what you see with what you know?
How can a ‘concept’ be ‘perceptual’ in nature?
How are such concepts learned, given that every language has a different set?
The first major insight came from Len Talmy, again in the summer of 1975. Talmy, looking at a wide variety of languages, made a major discovery. Spatial relations terms can be broken down into primitive spatial relations, where each language uses the same primitives, but puts them together in different ways. The central sense of English ‘on’, for example, uses the primitives Above, Contact, and Support. Not all languages have a complex concept corresponding to ‘on’, but they all have ways of expressing Above, Contact, and Support. Or take English ‘into’, which combines in and to in the following way: ‘In’ is defined relative to a Bounded Region schema; it locates an object at the interior of a bounded region. ‘To’ is defined relative to a dynamic Source–Path–Goal schema, and locates an entity on a path with a goal. ‘Into’ combines both the Bounded Region and Source–Path–Goal schemas, so that the goal is in the interior of the bounded region and the source is outside the bounded region.

Talmy also noticed that primitive image schemas fall into three types:

1. **Topological** (where relative nearness is preserved under deformations); Bounded Region is one example and a Path is another. Change their size and twist them around, and they remain bounded regions and paths.
2. **Orientational** (defined relative to bodily orientations); ‘in front of’ is an example.
3. **Force-dynamic** (making use of some kind of force); ‘against’ is an example.

In addition, each primary image schema comes with a ‘trajectory’ and a ‘landmark’. For example, in ‘The car is in the garage’, the garage is the landmark, relative to which the car (the trajectory) is located. Further detailed research on this topic has been done by Langacker (1986, 1990, 1991), Herskovitz (1986), Brugman (1981), and others.

What is important for our purposes is that all of these are embodied, with orientations like ‘in front of’ defined relative to beings with fronts, and force-dynamic schemas defined relative to how muscles operate.

Talmy’s insights led to decompositions of spatial relations into primitive image schemas with topological, orientation, and force-dynamic properties (see Talmy 2000). But Talmy and other cognitive linguists stopped short of explaining how image schemas
could serve both a perceptual and conceptual function, and failed as well to answer the question: Just what is an image schema anyway?

Those questions were answered by Terry Regier in his celebrated book *The Human Semantic Potential* (1996). Regier realized that certain well-known neural structures and mechanisms in the visual cortex have the right structure to ‘compute’ image schemas with visual properties. He reasoned, for example, that topographic maps of the visual field could compute topological image schemas (e.g., Bounded Regions and Paths) and that orientation-sensitive cell-assemblies could compute orientational image schemas (e.g., ‘above’ and ‘in front of’). Using techniques of computational neural modelling, Regier constructed a program that learns the spatial relation terms of a language and correctly names spatial relations projected on a computational model of the retina. In short, Regier came up with the only plausible hypothesis to date as to how image schemas are neurally embodied.

Regier’s model shows us how it is possible for very specialized neural structures in the visual system to ‘compute’ primitive spatial relations that are simultaneously perceptual and conceptual. The model explains how it is possible for all human beings to have the same primitive image schemas, which can be combined differently in different languages. It shows us how portions of the brain that evolved earlier can be adapted to characterize concepts. Finally it shows us just how dependent our concepts are on our peculiar neuroanatomy, which we inherited through evolution. Concepts did not just arise *de novo* with human beings. Human concepts require the use of the products of nonhuman evolution, adapting aspects of the sensory motor system to a new conceptual use. In order to think using spatial relations concepts, you have to have the right kind of visual system!

**Action Concepts and Verbs**

Regier had shown for the first time just how it is possible for a physical system – networks of neurons in the brain – to characterize concepts. It was the first step in answering the question: How can a brain, which works by neurochemistry, have ideas? Regier’s discovery began research in earnest on a neural theory of thought and language, both at Berkeley and at the University of Chicago,
where Regier took up a professorship. At Berkeley, the next dramatic advance was made by David Bailey.

Bailey asked how it was possible to learn words for embodied concepts. He limited his study to verbs of hand motion. Each language has such verbs—English, for example, has ‘push’, ‘pull’, ‘grasp’, ‘hold’, ‘drop’, and so on. The verbs in each language cover a different range of actions. But the range depends on what the hands can do. Movement, however, is complicated. Simple movements, like opening and closing the hand or pointing the index finger, are the products of motor synergies, complex circuits that are governed by small clusters of neurons in the motor cortex. Complicated movements are orchestrations of simple motor synergies performed by the prefrontal cortex, with neural connections called ‘bindings’ to the synergies in the motor cortex. When you move, a prefrontal circuit fires in sequence, coordinating the simple motor synergies dynamically and with feedback. How do you learn to connect verbs to such complex circuitry in just the right way?

The crucial insight came from an unlikely place—studies of how cats move. Cats have three gaits—a strut, a trot, and a gallop. Each is governed by very different neural circuitry. But, remarkably enough, the activation of these circuits is governed by a single neural cluster. If it is firing slowly, the cat struts; if it is firing at a moderate rate, the cat trots; and if at a high rate, the cat gallops. This is called a single neural parameter with three values.

For his research, Bailey needed to find a body—an online virtual body. He found one on the University of Pennsylvania website. Its name was Jack. It had every virtual bone, muscle, and motor synergy. All it needed to ‘move’ were the complex motor schemas with bindings to the motor synergies. Bailey needed to construct computational neural models of the motor schemas and bindings, and to model neural parameterizations, that is, minimal information structures, that could control the schemas. Then he designed a neural learning model that could learn to pair the right verbs with the right parameters and parameter values and to generalize correctly. When completed, Bailey’s model had learned both to provide the right verb given an action and the right action given a verb.

It was a remarkable accomplishment, but with an important theoretical bonus. We know from studies of mental imagery and dreaming that when you imagine or dream that you are moving your body or seeing something, the same part of the brain involved in real moving and real seeing is activated. The
difference is that, with dreams of movement, neural connections to the body are inhibited, and with visual imagination, inputs to the visual system come from the brain, not through the retina. Understanding what it means to move and see is thus a matter of imagination, the creation of a neural enactment using the same parts of the brain used to actually move and perceive.

The distinction between parameters and actual neural enactments has allowed for a Neural Theory of Language. In the theory now under development, word meaning and meaningful grammatical constructions are characterized via the pairing of neural parameters for embodied concepts with neural parameters for phonetics, linear order, and other aspects of linguistic form. The result is an embodied theory of language understanding.

But what about intentionality – the link between concepts and the world? What about abstract thought – about concepts like justice, love, morality, and causation? What about inference? What could it mean for inference to be embodied? And how is abstract reasoning embodied? What could it even mean for abstract concepts and reasoning to be embodied? To answer these questions, we need first to return to the late 1970s.

**Conceptual Metaphor**

Perhaps the most radical discovery that changed the theory of mind was made by Michael Reddy in 1977 and independently by me in 1978. Reddy, in his classic paper on the Conduit Metaphor (1979), showed that the very concept of communication is metaphorical. Not metaphorical according to Aristotle’s theory of metaphor, which most of us learned in school. It turns out Aristotle was mistaken. What most of us learned in school as a ‘definition’ of metaphor is in fact a false theory.

Metaphor is not a mere matter of words, not based on similarity, not just a feature of poetic or rhetorical language, and not deviant. Rather, metaphor is conceptual, not merely linguistic. A metaphor is a systematic conceptual mapping from one conceptual domain (the source) onto another (the target). It may introduce conceptual structure. And metaphor functions primarily to allow sensory-motor reasoning to apply to subjective judgements. No concept is wholly metaphorical; there is usually some minimal conceptual skeleton in the target domain – though not enough to do much reasoning with. Most of the abstract inferences—and much of the lexicography—for abstract concepts come via metaphor.
Reddy's original example of the Conduit Metaphor is a good example. The mapping looks like this:

- Ideas Are Objects.
- Words Are Containers (for idea-objects).
- Communication Is Sending (idea-objects in word-containers).

In the metaphor, a communicator puts idea-objects into word-containers and attempts to 'get the idea across' to her interlocutor. Communication is successful if her interlocutor 'gets' what she says. The metaphor has further details. Idea-objects don't fit into any old word-containers; there are right and wrong words for ideas, and it is up to the speaker to put her ideas in the right words. In most cases, 'the meaning is in the words'. But when a speaker is insincere, her words may be 'hollow' or 'empty'. A speaker who is trying not to communicate directly may 'hide her meaning' in 'dense' paragraphs. Reddy lists more than 140 such common, everyday metaphorical expressions for this one conceptual metaphor.

Eve Sweetser (1990) and Alan Schwartz (unpublished) have observed that the Conduit Metaphor is a special case of a much more general and elaborate metaphor system – The Mind As Body system (see Lakoff & Johnson 1999: ch. 12). The general mapping is as follows:

- The Mind Is A Body.
- Thinking Is Physical Functioning.
- Ideas Are Entities (relative to which the body functions).

This general metaphor has four special cases of physical functioning: Manipulating objects, Perceiving, Moving, and Eating.

The Conduit Metaphor is a special case of Thinking As Manipulating Objects. This metaphor includes the conception of Understanding As Grasping, Teaching As Providing Students With Ideas, and includes such expressions as 'tossing ideas around', 'playing with ideas', and 'shaping a theory'.

In the special case where Thinking Is Perceiving, there are the following metaphors: Knowing Is Seeing, Coming To Know Is Observing, Understanding Is Seeing Clearly, Communicating Is Showing, and so on. This metaphor is used in expressions like: 'shedding light on the subject', 'being enlightened', 'pointing out a fact', 'a clear presentation', a murky paragraph', and so on.
Another special case is Thinking Is Moving, in which one can ‘lead someone step by step through an argument’, ‘follow an argument’ or ‘get lost’, ‘talk in circles’, ‘go directly to the point’, ‘reach conclusions’, ‘skip steps in an argument’, or ‘zoom through a lecture’.

My favourite is Thinking Is Eating, where Ideas Are Food, Communicating Is Feeding, Accepting Is Swallowing, Understanding Is Digesting, and so on. This metaphor gives rise to such expressions as ‘spoon-feeding your students’, ‘regurgitating information in the exam’, ‘letting ideas simmer for a while’, and so on. French, as you might expect, has a very elaborate version of this metaphor. My favourite expression is ‘aux petits oignons’ (stewed with spring onions), which means a particularly exquisite idea.

These examples are clearly metaphorical. They are systematic. They involve applying the reasoning of the physical source domains to the abstract target domain. Moreover, they define a huge proportion of our modes of understanding what ideas, thought, understanding, and communication are. Try having a conversation about thinking, communicating, and understanding for ten minutes without using any of these metaphors or any of the reasoning that arises from their use. I bet you can’t. Note that even when you speak of communicating an idea to someone, the to is coming from the directional path of the Conduit Metaphor, where Communication Is Sending an idea to someone.

Three other important metaphors for ideas are these:

• Thought Is Language, with examples like ‘Do I have to spell it out for you?’, ‘Let me make a mental note of that’, ‘She’s an open book to me’, ‘I can read her mind’, ‘The argument is abbreviated’, ‘He’s reading between the lines’, ‘That’s Greek to me’, and so on.

• Thought Is Mathematical Calculation, with expressions like ‘It doesn’t add up’, ‘What does it all add up to?’, ‘What’s the bottom line?’, ‘Give me an accounting of what went on’, and ‘We won’t count that’. And finally:

• The Mind Is A Machine, including such expressions as ‘I’m feeling a little rusty today’, ‘The wheels are really turning now’, ‘He’s cranking out ideas’, and ‘He had a breakdown’.

Each of these metaphors for mind conceptualizes ideas in a somewhat different way, with different inferences. Yet they define the normal way we think of ideas and of the mind. One
of the most interesting developments to come out of the theory of conceptual metaphor is the discovery of the mechanism of conceptual blending, by Gilles Fauconnier and Mark Turner (2002). In a blend, distinct conceptual structures, including metaphors, can be bound together in well-specified ways. A particularly important blend is the complex metaphor for mind that binds together the metaphors for thought as step-by-step movement, as the product of a machine, as mathematical calculation, as language, and as object manipulation. The result is the central metaphor of Artificial Intelligence, namely, that The Mind Is A Computer, a machine that thinks by doing mathematical calculation via manipulating symbols in a 'language' step-by-step. Because each of the component metaphors is a commonplace part of our conceptual system, it is easy to think of the mind as a computer. This metaphor was also the basis of first-generation cognitive science back in the 1960s and 1970s, and it is still used today.

This use of these metaphors is actually part of a more general blend – one that characterizes a major philosophical movement: Anglo-American analytic philosophy. There are many variants of Anglo-American analytic philosophy, for the most part they include certain basic tenets:

- Thought can be represented adequately without reference to the body.
- It can be modelled as the manipulation of symbols that are meaningless in themselves.
- Truth is correspondence between symbols (either words or representations of concepts) and the world.
- Meaning is literal and based on truth. Meaning is public, and shared.
- Concepts are defined as a set of necessary and sufficient conditions expressed symbolically.
- Philosophy consists in conceptual analysis.

These tenets are entailments of the set of everyday metaphors for mind that we have just described.

We can see this by looking at a collection of 17 entailments of the metaphors in this system. Entailment (1) is part of the blend because in each case ideas are mapped onto external objects in the world that exist independent of the speaker: physical objects, locations, food, linguistic symbols, numbers, and products of machines.
The Mind As Body System
(1) Ideas have a public, objective existence independent of any thinker.
(2) Ideas correspond to things in the world.

Thought As Motion
(3) Rational thought is direct, deliberate, and step-by-step.

Thought As Object Manipulation
(4) Thinking is object manipulation.
(5) Ideas are objective. Hence, they are the same for everyone, that is, they are universal.
(6) Communicating is sending.
(7) The structure of an idea is the structure of an object.
(8) Analysing ideas is taking apart objects.

Thought As Language
(9) Thought has the properties of language.
(10) Thought is external and public.
(11) The structure of thought is accurately representable as a linear sequence of written symbols of the sort that constitute a written language.
(12) Every idea is expressible in language.

Thought As Mathematical Calculation
(13) Just as numbers can be accurately represented by sequences of written symbols, so thoughts can be adequately represented by sequences of written symbols.
(14) Just as mathematical calculation is mechanical (that is, algorithmic), so thought is.
(15) Just as there are systematic universal principles of mathematical calculation that work step-by-step, so there are systematic universal principles of reason that work step-by-step.
(16) Just as numbers and mathematics are universal, so thoughts and reason are universal.

The Mind As Machine
(17) Each complex thought has a structure imposed by mechanically putting together simple thoughts in a regular, describable, step-by-step fashion.

These 17 entailments of our everyday metaphor system for the mind characterize the central tenets of Anglo-American analytic
philosophy, as Johnson and I discuss in great detail in *Philosophy in the Flesh* (1999: chs 12 and 21). That we already have these metaphors makes such a philosophy seem natural to many English speakers (though certainly not to all!).

One of the ironies about this is that Anglo-American analytic philosophy, even though it is defined by this combination of metaphorical entailments, does not—and cannot—recognize the very existence of conceptual metaphor. The reason is that concepts, in that philosophical theory, have to correspond to things in the world that exist independent of any thinker. This is a version of the correspondence theory of truth. Obviously, conceptual metaphors are aspects of human cognition. The correspondence theory of truth cannot hold for them. The idea that concepts are concepts of things in the world—to be defined by necessary and sufficient conditions on external reality—cannot hold for conceptual metaphors. Thus, there is a contradiction between the existence of conceptual metaphors and Anglo-American philosophy. It is no wonder then that conceptual metaphor has not found a place within Anglo-American analytic philosophy. This is especially true of formal semantics, where meanings of concepts are seen as mapping onto set-theoretical models, taken as models of an external reality assumed to consist of objects with properties and relations. Conceptual metaphors just don’t fit that philosophically driven theory of semantics.

And yet, the irony is that the central tenets of the philosophy arise from our most commonplace conceptual metaphors about the mind and ideas!

### Inference, Embodiment, and Intentionality

Our trip back into the 1970s gave us a glimpse into how abstract concepts and abstract reason could be embodied, namely, via conceptual metaphors with sensory-motor source domains! That’s a bit oversimplified as we shall see, but it’s a good approximation for now.

Our story now resumes in Berkeley in the mid-1990s. As David Bailey was modelling the learning of verbs of hand motion, he recruited the help of the student in the office next door, Srini Narayanan. Narayanan created a precise computational means of modelling all the relevant neural information needed for Bailey’s task. What he did was take Petri nets, a common off-the-shelf form
of AI, and extend them in important ways to do appropriately restricted neural modelling.

Narayanan then took 40 or 50 of Bailey’s motor schemas and represented them in his model. As he did so, he made a remarkable discovery: *They all had the same structure!* Recall that motor schemas come in three parts: the simple motor synergies controlled by the motor cortex; the complex motor schemas in the prefrontal cortex; and bindings between the two. When the synergies and the bindings were factored out because they occurred elsewhere, what remained for prefrontal motor control were neural schemas of the same basic structure – with various options and alternatives.

That structure is simple:

1. getting into a state of readiness;
2. the initial state;
3. the starting process;
4. the main process (either instantaneous or prolonged);
5. an option to stop;
6. an option to resume;
7. an option to iterate or continue the main process;
8. a check to see if a goal has been met;
9. the finishing process;
10. the final state.

This should come as no surprise. Any high-level motor activity you undertake, from scratching your head to turning on a light switch to sipping a cup of tea, will have this structure. There are variations and complications, of course. One of these structures can be embedded in another. The structure may be encapsulated as a single simple element. There are options: purpose or no purpose, iteration or no iteration, an instantaneous or extended process, encapsulation or no encapsulation, and so on.

Narayanan presented his results at the research group run by Jerome Feldman and myself at the International Computer Science Institute. I immediately recognized that structure, as any linguist would. It is the structure of aspect in the languages of the world, the structure that linguistic systems impose to structure events. Verbs have inherent aspects: the verb ‘tap’ is inherently iterative and imperfective. Inherently ‘tap’ does not have a final state built in, while ‘open’ does – the state in which the object opened has become accessible, the ‘open’ state. Every language also has a collection of grammatical and/or morphological
means to change the inherent aspect of the verb. 'Is opening' picks out the central process, 'has opened' designates a final state, and 'is about to open' designates the state of readiness.

But aspect applies to any event at all, not just bodily movements. If aspect has the same structure as all motor control, that means that we understand the structure of any event at all (no matter how 'abstract') in terms of what our bodies can do. Not only is this not strange, but there is a well-known conceptual metaphor, Events Are Actions, by which any event can be conceptualized as an action by an agent with a body. An example is 'The wind blew open the door', where the wind is conceptualized metaphorically as a living being, blowing on the door.

Importantly, aspect has a logic that is difficult to characterize and which had been worked on extensively but never fully worked out. Narayanan then asked a fateful question. Could the same neural system that could move your body also perform abstract inferences about events? Could it reason that, after France fell into a recession, it was in a recession, that the fall was not intentional, and so on? In other words, could the general neural motor control schema that he had found also do abstract reasoning? Could the system that controls bodily movements also reason about abstract domains that are metaphorically conceptualized in terms of bodily movements?

To test this, Narayanan turned to the theory of conceptual metaphor, where a mass of results has indicated that abstract reasoning is done in terms of sensory-motor reasoning. Narayanan took as his abstract domain the field of international economics. He then did an Internet search of the Wall Street Journal, the Economist, and the New York Times Business Section looking for examples of verbs expressing physical actions, for which he had a model of the neural control structure for the action. He came up with thousands of examples and chose 30 for his dissertation. These were cases such as ‘France fell into a recession; pulled out by Germany’, ‘India stumbles in liberalization plan’.

Narayanan then constructed a neural theory of metaphor, in which conceptual metaphors are neural circuits linking source and target domains. His model contained three fundamental parts:

1. Simple circuitry characterizing the literal logical relations holding among economic concepts. These of course could not provide the right inferences for understanding the metaphorical sentences in his news stories.
2 Motor control circuitry for performing the bodily actions in the source domains.

3 Metaphorical mapping circuitry, so that the motor inferences could be mapped onto abstract economic concepts. These mappings allowed the motor inferences to combine, via metaphor, to achieve abstract reasoning that could be combined with the purely abstract economic inferences.

The model worked. It successfully characterized the inferences from the new stories. The same neural circuitry that could run a body could do abstract inference!

That doesn’t mean that the brain actually works this way, but it is suggestive. The brain tends to use structure already developed and adapt it to other tasks. If it has structures that are being used for complex motor control, and if it has a neural mechanism for conceptual metaphor, linking the motor system to higher cortex, it would make sense if the motor control system together with the existing metaphorical connections actually performed the abstract reasoning rather than have essentially the same circuitry copied in a very different part of the brain.

In fact, one should be able to test this hypothesis using fMRI technology. The testing has already begun. Tim Rohrer (2001) at the UCSD cognitive science department, working with Martin Sereno and Marta Kutas, has reported a positive result in initial studies. Touching the hand activates a well-known region of motor cortex. Literal language for the hand – e.g. ‘grasp’, ‘manipulate’ – activates a subpart of the hand region of the motor cortex. And so does metaphorical language about the hand, for example: ‘Can you grasp what I’m saying?’ ‘Don’t try to manipulate me into accepting your ideas’. The first results indeed point in the direction of confirming Narayanan’s hypothesis: The same neural circuitry that can move your body can do abstract inference via conceptual metaphor.

Incidentally, this evidence is the tenth type of evidence to be used to substantiate the existence of conceptual metaphor. A survey of experimental evidence from cognitive psychology appears in The Poetics of Mind by Raymond Gibbs (1994). Impressive recent experiments have been performed by Lera Boroditsky of the MIT cognitive and brain sciences department (2000). A survey of other evidence types is given in Lakoff and Johnson (1999: ch. 6). At this point, there can be little doubt that conceptual metaphor is cognitively real.
Primary Metaphor and the Neural Theory

In *Metaphors We Live By* (1980), Mark Johnson and I hypothesized that certain very basic conceptual metaphors arose from correlations in everyday experience. The example we gave was More Is Up, with expressions like ‘The temperature is rising’, ‘Stock prices hit bottom’, ‘Thefts have soared in London’, and so on. We suggested that the regular correlation of quantity with verticality, as when one creates piles or pours liquid into a glass, would result in such a metaphor.

In 1997, this hypothesis was confirmed in child language acquisition research by Christopher Johnson. In a study of the acquisition of the Knowing Is Seeing metaphor, Johnson found that children first learn literal ‘see’ as in ‘See doggie’ and ‘See Daddy’. Then they learn cases which he referred to as ‘conflations’, where the domains of seeing and knowing are co-active, that is, where both are at issue, as in sentences like: ‘See Daddy come in. See what I spilled’. Finally, children learn pure metaphorical cases such as ‘See what I mean’. Johnson has argued that metaphor arises from such conflation, or neural co-activation, in everyday experience of the source and target domains of the metaphor. The cases he looked at occurred before the age of three.

Narayan’s neural model of metaphorical mappings comes with a learning theory, called ‘recruitment learning’, which explains Johnson’s results. In recruitment learning, when neural clusters are regularly co-active, neural connections of a certain type are learned between them. As the slogan goes, ‘Neurons that fire together wire together’. Narayan’s theory predicts that large numbers of simple conceptual metaphors arising from conflations should be learned in childhood, beginning before the age of three and extending after that.

Joseph Grady, also in 1997, argued on independent grounds that such was indeed the case. He has found persuasive linguistic evidence that complex metaphors of the sort described by Johnson and myself should actually be described as conceptual blends of simple metaphors that arise by conflation, what he calls ‘primary metaphors’. These are well-known cases like States Are Locations, Purposes Are Destinations, Causes Are Forces, Anger Is Heat, and so on. Grady, in his 1997 dissertation, examines a wide range of such cases, pointing out that such primary metaphors appear to be widespread around the world, since most human beings experiences such conflations regularly. For
example, if you want to achieve a *purpose*, it is common to have to go to a particular *destination* to do it.

Grady’s research gave substance to Johnson’s conflation research and Narayanan’s neural recruitment model, since it showed that apparent counterexamples (apparently ungrounded complex metaphors) were actually blends and extensions of primary metaphors.

Taken together, the research by Narayanan, C. Johnson, and Grady jointly provides a neural theory of the acquisition of conceptual metaphors. The prediction is that in early childhood, say by the age of four, we should have learned hundreds of conceptual metaphors (not all of them linguistically instantiated). These should guide the formation of new abstract concepts and forms of abstract reason based on sensory-motor inference.

**Mirror Neurons**

The last part of our story is perhaps the most interesting. We shift scenes from Berkeley to Parma, Italy, where a group of neuroscientists led by Giovanni Rizzolatti and Vittorio Gallese made a remarkable discovery (Rizzolatti *et al.* 1996; Rizzolatti & Arbib 1998; Gallese *et al.* 1996; Fadiga *et al.* 1995). They discovered, in a study of macaque monkeys, that the same neurons in the prefrontal cortex that fire during complex motor actions also fire when the animal sees a person or another monkey doing the *same* motor action. In short, the neurons in this part of the prefrontal cortex are linked both to visual perception and motor action, *for the same motor actions*. These are called ‘mirror neurons’. In a region adjacent to the mirror neurons, they found what they called ‘canonical neurons’. These fire either when a complex motor action is performed *or when the animal sees an object of the kind that the action is normally performed on*. The same canonical neurons that fire when the monkey eats a banana fire when the monkey sees the banana. The mirror neurons, by contrast, fire when either the monkey eats the banana or sees another monkey eating a banana, but not just when the monkey sees a banana. This research has since been extended to human beings (Iacoboni *et al.* 1999).

Think back now to the Bailey and Narayanan research pairing verbs with neural parameters governing actions. Action verbs apply not only to one’s own actions, but to the actions of others
as well. But Narayanan’s models were originally created only for one’s own action, not actions by others. At that point, it was a problem to relate first-person experience with third-person actions. The discovery of mirror neurons linked first-person experience to third-person descriptions, since the same neural executing schema (or ‘X-schema’) over mirror neurons could characterize both.

Here is how this idea was seen as working technically: The neural X-schemas that Narayanan was modelling have a neural structure that can be characterized independently of its bindings to the motor synergies. Such a structure can also have bindings to schemas for visual perception as well. In short, Narayanan’s models are just what would be appropriate for modelling structures over the mirror neurons and canonical neurons in human beings. With metaphorical connections added, such embodied structures can characterize abstract reasoning.

Up to now we have characterized a theory of understanding in terms of neural enactments of sensory-motor actions and perceptions, together with any metaphorical neural mappings to abstract concepts. This account left out intentionality. Mirror neurons allow us to include it. The mirror and canonical neurons connect our brains to the world we function in with our bodies. They allow us to characterize a neural theory of mind that includes intentionality – the connection of actions to objects via canonical neurons.

As we shall see in Chapter 4, mirror neurons do even more than that. The region of the brain where they occur is connected to the brain’s emotional centres. It appears that such circuitry governs empathy. We know from Paul Eckman’s research that configurations of facial muscles express certain emotions. Presumably, our mirror neurons fire when we see the same configurations of facial muscles on someone else that our facial muscles would make. And that firing can activate our own emotional centres. In short, that allows us to empathize – to feel someone else’s pain or joy. We will see in Chapter 4 how this relates to an embodied moral theory. The mirror neurons and canonical neurons allow our brains to go outside ourselves. Chapter 4 will also focus on what that has to do with an embodied account of religion and spiritual experience.
Grammar from a Neural Perspective

I am a linguist, and so it should not be surprising that I am concerned with grammar. From the perspective of the Neural Theory of Language, as it is being developed at Berkeley, grammar turns out to be a system of learned neural connections between the neural parameters of our conceptual system (which is grounded in the sensory-motor system) and the neural parameters of our phonological system (also grounded in that part of the sensory-motor system concerned with phonetics or signed languages). In short, grammar consists of learned neural connections between the independently grounded conceptual and phonological systems. Grammatical constructions are thus not innate. Grammar is inherently cross-modal, rather than modular. And grammatical universals follow from (1) shared aspects of embodied conceptual systems, (2) the fact that grammar lies in conceptual–phonological links, (3) the properties of neural systems carrying out those linkages, and (4) in the way in which grammar is used in communication. This is the very opposite of what Chomsky had proposed back in the 1950s on an a priori philosophical basis.

Conclusion

The embodied metaphorical mind is a new account of mind. It comes out of research in neuroscience, cognitive linguistics, and other branches of cognitive science. It contradicts earlier philosophically based cognitive theories. It happens to share certain features with the ideas of philosophers like Merleau-Ponty and Dewey, but goes well beyond what they had envisioned. It contradicts the basic tenets of Anglo-American analytic philosophy and most of twentieth-century Continental philosophy as well.

The new view of mind calls for a new philosophy, what Mark Johnson and I have called, from the perspective of issues like truth and meaning, 'embodied realism', and what we have called 'experientialism' from the perspective of issues like morality, aesthetics, spiritual experience, and all issues pertaining to how to live one's life. Those are the issues I will take up in the next chapter, beginning with questions of consciousness, causation, and the nature of mathematics.
How we understand the nature of mind is all-important for a wide range of very general human concerns. Given the account of the embodied mind in Chapter 3, we are in a position to see just how important it is to understand the nature of the mind. The issues I discuss below are consciousness, the soul (if there is one), causation, the nature of mathematics and mathematical truth, why mathematics works in science, and finally morality and religion. Since I have discussed many of these matters elsewhere, I will be stressing the overall relationship among these topics within an experientialist philosophy, rather than the details of each.
The Problem of Consciousness

What has been called 'the problem of consciousness' has many dimensions. In Chapter 3, I discussed only one of those dimensions, the Neural Language Problem: How the physical brain gives rise to ideas and language – how detailed brain structures and mechanisms result in the detailed conceptual and linguistic structures that we use in drawing inferences and communicating.

What has been called 'consciousness' is not a single thing, but the confluence of many distinct phenomena. It is made up of at least the following aspects, which may well have distinct characteristics and different origins.

1. **Neural language**: The neural computational aspects of thought and language. Though unconscious in themselves, these underlie and make possible whatever conscious thought and language use occurs.

2. **Attention**: The capacity to focus on one aspect of experience rather than another.

3. **Memory**: Not only the ability to remember events, but also the capacity to use previously developed sensory-motor patterns in perception and action, what Edelman (1992) refers to as 'the remembered present'.

4. **Self-monitoring**: The ability to monitor what one is doing and thinking.

5. **First-person perspective**: The perspective of an experiencer or actor.

6. **Unity**: We know our brains compute, say, colour and shape in different places, yet we experience the colour and shape of objects in a unified way. The problem of the unity of consciousness is the problem of how the many parts of our brains produce a unified sense of experience.

7. **Mental causality**: The ability to decide to perform some action and consciously carry it out, monitoring it as one does it.

8. **Self-awareness**: The fact of sentience, of sensing what is happening.

9. **Qualitative experience, or qualia**: The way things look, sound and feel – why red looks red, not green; how a cello sounds; how a rose petal feels.

All of these aspects of consciousness are under intense study, and extraordinarily interesting research is being done on many of
them. (See, especially, Crick 1994; Damasio 1999; and Edelman & Tononi 2001.)

It is an open question to what extent these phenomena are amenable to scientific study via the methods of neural computation. Francis Crick, for example, has argued plausibly that attention can be studied by such means (Crick 1994). I believe that all but two of the dimensions of consciousness can ultimately find an explanation in neural computational models. Those two, which stand outside of neural computation, are awareness and qualitative experience, or qualia. I can imagine no way in which these two aspects of consciousness could be characterized in terms of current neural computation.

Although I believe that there can be no complete neural computational theory of consciousness, there are correlations between aspects of consciousness and the neural underpinnings of reasoning and language use. For example, our brains create colour on the basis of neural impulses transmitted by colour cones in the retina. A great deal is known about the neural circuitry involved in colour perception. But the computational properties of those circuits do not, and cannot, account for the way that red looks and how it differs from the way that green looks. Nonetheless, there is a correlation between the levels of firing of certain neurons and our qualitative experience of colour. But we know that correlation is not causation. There is no causal model of how the firing of certain neurons results in our qualitative experience of red, while the firing of other neurons results in our qualitative experience of green.

I mention this for an important reason. Most people think that their thought processes are conscious, while research in cognitive science has shown that the mechanisms of thought, like the mechanisms of language, are mostly unconscious. Indeed, the theory we will be presenting here is a theory of unconscious conceptual and linguistic structures and processes. Our claim is that the mechanisms of thought and language are neural computational mechanisms, of which we are unaware – that is, mechanisms we are not capable of being conscious of. But if thought is unconscious, why do we have the feeling that it is conscious?

The answer, I believe, lies in the correlations between certain kinds of neural activity and corresponding qualitative experience. Take, for example, the qualitative experience of hearing sounds. A considerable amount is known about the neural activity in the auditory cortex. This neural activity correlates with something we do not understand – the qualitative experience of hearing
sounds, such as speech sounds. The kind of neural computational theory given here could, in principle, give an account of the kinds of neural computations going on in the auditory cortex when we hear speech sounds. Somehow or other, in ways we do not understand, there is a correlation between neural activity and the qualitative – conscious – experience of hearing.

Now consider the experience of ‘hearing yourself think’ when no sound is actually being produced. How can you ‘hear’ when there is no sound? Yet we all do.

Hearing involves the activation of neurons in the auditory cortex – usually by neural connections coming from the ear. But neural activation in the auditory cortex need not originate in the ears. The activation could originate elsewhere in the brain. Imagine hearing someone mention your name. You can ‘hear’ your name being said, even though there is no sound. This is a product of auditory imagination. The same is true of dreaming. You can ‘hear’ sounds in your dreams, even though no physical sound was ever made. Auditory imagination is, of course, considerably degraded; it doesn’t sound like the real thing. Yet it is clearly recognizable as an instance of the real thing.

The auditory cortex takes input not only from the ears, but also from other areas of the brain. When you speak, you are thinking using concepts expressible in language. The neural circuitry constituting the grammar of your language links the areas of the brain concerned with thought with those areas devoted to phonetics. The activation in the regions of thought flows via those grammatical circuits to the phonetic areas (speech and hearing). If we speak, we move our mouths and produce sounds. But we can also imagine speaking. Imagine saying the sentence, ‘How are you?’ When you do, you can ‘hear’ what you are imagining yourself saying. The reason appears to be that, as with the visual cortex, your auditory cortex is getting input from other areas of your brain. You are ‘hearing yourself think’ – that is, ‘hearing’ the words expressing your thoughts.

There is a qualitative experience in hearing yourself think. You can ‘hear’ differences in vowel quality: in your mind, say ‘ah’ to yourself and then say ‘eeh’. You can ‘hear’ the qualitative differences among speech sounds – even though there is no sound. The reason is that the some of the same parts of the auditory cortex are being activated as when you actually hear the sounds with your ears. When you hear yourself think, the conscious qualitative experience of ‘hearing’ correlates with activation of the auditory cortex. The activation of the cortex can be modelled by
neural computation; the conscious qualitative experience, which is a very different kind of thing, cannot.

I believe that the feeling we have that thought is conscious comes in large part from the conscious qualitative experience of 'hearing' activations in the auditory cortex produced when we think. It also comes from other types of conscious qualitative experience that is triggered by and correlates with thought, for example, mental imagery or motor imagery.

But we know from a wide variety of experimental evidence that thought itself – the activation of concepts and the performance of inferences – is largely unconscious, as is the activation of the corresponding phonological means to express thought. I believe, following central results in cognitive science, that thinking and the mapping of thought to the appropriate linguistic forms are unconscious and achieved via neural computation, as discussed in Chapter 3.

The Computation–Consciousness Correlation

The basic mechanisms of thought – conceptual structure, mental simulation, and inference – can be modelled using neural computation. What are beyond neural computation are awareness and qualitative experience, or qualia. I will call these, jointly, qualitative awareness. Qualitative awareness seems to occur only when there is corresponding neural activation, which can be modelled via neural computation. In short, neural activation may not be able to model qualitative awareness, but it can model when that awareness occurs. For example, a sequence of thought can be modelled via neural computation. The corresponding stream of consciousness cannot be modelled by neural computation, but the ordering of the corresponding neural activations can be. The same is true of inference. Inferences can be modelled by neural computation at the neural level. Our qualitative awareness of inference cannot be modelled by neural computation, though the ordering of the corresponding neural activations can be.

This means that neural computation is anything but irrelevant to consciousness. The type of qualitative awareness you have is a matter of neural computation, though qualitative awareness itself is beyond neural computation. Since neural computation can (in principle) model when and where neural activation occurs, and since neural activation is a prerequisite for qualitative awareness,
it follows that neural computation is necessary but not sufficient for an account of qualitative awareness, and hence of consciousness.

The tight correlation between qualitative awareness and neural activity suggests strongly that consciousness is embodied – neurally embodied, even though the mechanism of consciousness is not neural computation alone.

The Soul

The soul is usually seen as the locus of consciousness. Consciousness appears to be embodied. But the soul, since it is supposed to live on after death, must be disembodied. There is a contradiction here.

Cognitive science cannot tell us whether or not souls exist – whether there is life for a soul after the body dies, or before it is born, as in theories of reincarnation. But the cognitive science of the embodied mind can tell us what kinds of experiences you need a body for. If there is a soul that lives after the body dies, then it cannot have those experiences and perform those actions. In short, cognitive science can tell us a lot about what a soul isn’t, if souls exist at all. Suppose, for the sake of argument, we assume that souls do exist. Then we can ask, as a scientific question, ‘What is a body required for?’ And hence, ‘What can a disembodied soul not do?’

The general argument has the following form:

If certain peculiarities of thought and experience necessarily result from corresponding peculiarities of the way we are embodied, then they cannot be attributed to a soul conceived of as disembodied.

For example, consider the question of whether the soul can see colour. Without a body, the soul has no colour cones and none of the requisite circuitry to compute colours. Since colours are not out there in the world, and since they require these parts of a body, it follows that the soul by itself cannot compute colours.

Well, what about vision? Can a soul see at all? Let’s start with 3-D vision, which requires a pairing of neural inputs from the left and right retinas lined up in ocular dominance columns with corresponding left and right inputs next to each other. The information required for 3-D vision comes from the difference between left and right neural activity in such pairs. But without
where patients lose function in the right frontal and/or temporal lobes of their brains. Here are descriptions of what happens: Patients with asymmetric right frontal dysfunction exhibit marked behavioural alterations which include: verbal disinhibition, antisocial behaviour, loss of concern for others, and changes in previously established patterns of dress and political ideology. Those with selective right temporal dysfunction exhibit loss of empathy, intensification of political or religious ideas, verbal preoccupation, and blunting of emotional feelings.

Dr Bruce L. Miller, of the University of California at San Francisco, has described such cases as ‘changes of self’. Who you are has everything to do with how you behave towards others, whether you are empathetic or thoughtful, whether you are discreet, whether you care what others think of you, how you dress, what your political and religious beliefs are, what your emotional reactions are. When all these can change radically and permanently because of brain degeneration, it’s hard to say who ‘you’ really are: the ‘you’ before or after the brain change? If your personality and character can change because of a bodily change in the brain, and if your soul is unchanged when your body changes or dies, then the soul cannot be the locus of your personality and character. In other words, if there is no one thing that you essentially are independently of your body, then the soul could not be the locus of who you essentially are.

If we assume a soul exists and that it is disembodied, then it cannot have any of the properties that require a body. In short, the soul, if it exists:

- cannot see, hear, or otherwise perceive;
- cannot think using ordinary human concepts;
- cannot empathize with human beings;
- is not conscious;
- is not the locus of anyone’s personality or character.

But these are exactly the properties that most people take as definitional of a soul that lives on after death. What’s the point of having a ‘soul’ that lives on after death, but can’t perceive, think, empathize, have conscious experience, and doesn’t have your personality or character?
bodies, souls have no retinas and no corresponding neural circuitry, hence no 3-D vision.

You can see where this is going. Take any aspect of vision that is well understood via neuroscience, where certain sense organs and neural circuitry is required. Then a disembodied soul has no such bodily apparatus and so is incapable of that aspect of vision. Since a huge amount is known about the neuroscience of vision, it seems clear that the soul cannot see.

Hearing works the same way. For hearing you need an acoustic cortex, which consists of a frequency-intensity map getting input from sound-wave stimulation of the inner ear. With neither a cochlea nor an acoustic cortex, a soul cannot hear. Indeed, a soul cannot perceive at all.

But what about thinking? Could a soul think? As we saw, a body is necessary for human concepts and thought using those concepts. For spatial relations concepts, we need topographic maps of the visual field, orientation-sensitive cell assemblies, and much more. Souls could not have visual image schemas. Since a (left) prefrontal cortex is needed to characterize reasoning about events and actions, disembodied souls could not reason about events and actions. And if abstract thought mostly uses conceptual metaphors with source domains in the sensory-motor system, it follows that metaphorical thought is beyond souls. In short, reasoning as we know it using ordinary human concepts cannot be done by disembodied souls.

Could a soul empathize? As we saw in Chapter 3, empathy appears to require the mirror neurons and connections to the emotional centres of the brain. To recognize the feelings someone else has, you need to be able to see what their facial muscles are doing, sense via the mirror neurons what it would be like for your facial muscles to do that, and know, via connections to the emotional centres, what you would feel like if your face were doing that. The soul could not empathize without a body.

Could a soul be conscious? Would it have qualitative awareness, that is, awareness and qualitative experience? As we have just seen, qualitative awareness depends on, and correlates with, neural activity. With no neural activity, there is no qualitative awareness. In short, the soul, being disembodied, not only could not perceive or think, but it would have no consciousness at all.

Traditionally, it is believed that your soul, not your body, makes you who you are. Your soul characterizes your essence – your basic personality and character. However, in recent years, cases have been found of a disease called FTD, or frontotemporal disorder,
Causation

The idea of a disembodied soul as the seat of conscious thought is representative of a general denial of the body at the heart of Western culture, religion, and philosophy. The disembodied soul is sister to disembodied thought, disembodied knowledge, and disembodied truth. Once one understands the details of how thought, knowledge and truth are not just mediated by the body, but rather shaped by the body, our world view changes, not subtly, but massively. I turn next to causation, because it is central to issues of knowledge and truth, especially in the sciences.

One of the many great contributions of cognitive linguistics to cognitive sciences has been its concentration on the full range of conceptual and linguistic phenomena, phenomena often overlooked by cognitive psychology, philosophy, and studies in AI. Causation is a case in point. An important source of evidence comes from theories in the sciences and social sciences:

- **Causal paths and trees**: Change depends on other changes. An example is the QWERTY keyboard, first introduced in the days of the manual keyboard to keep typists from typing so fast that the keys would get stuck. Once millions of typists learned the keyboard, it was difficult to switch. Having started on the QWERTY path, there was no way to change.

- **The domino effect**: Once one country falls to communism, then the next will, and the next … until force is applied to keep one from falling (from Vietnam War days).

- **Thresholds**: For a while there is a build-up with no effect, but once change starts, it becomes uncontrollable.

- **The plate tectonic theory of international relations**: When causal force is applied to something large, the effect lags after the action of the cause. This is used to explain the fall of communism long after the USA started putting ‘pressure’ on communist governments.

These are real examples, taken from the social sciences, of metaphorical causal models. Each has its own logic, taken from some other domain. They show that causal models, as really used, do not share one and only one logic; rather, there are many logics of causation – and they are metaphorical.

This is true even more dramatically of the causal concepts that we use in everyday life. Many studies of causal concepts by psychologists or philosophers just look at uses of a few words,
such as 'cause' and maybe 'allow' and 'prevent'. The first cognitively serious study of causation to go beyond this was Leonard Talmey's classic paper, 'Force dynamics in language and thought' (1985), which first showed that causation is metaphorically based on our embodied use of force in everyday life. Causes Are Forces is a primary metaphor, learned automatically and unconsciously in early childhood. It lies at the centre of an elaborate metaphor system for causation, described in meticulous (if not excruciating) detail in Lakoff and Johnson's Philosophy in the Flesh (1999). Since I am a linguist, I'll start where linguists usually start, with some representative sentences exemplifying causation. The words in bold express causation — of one sort or another.

The noise gave me a headache.
The aspirin took it away.
The democrats blocked the balanced budget amendment in the senate.
FDR's leadership brought the country out of the depression.
The home run threw the crowd into a frenzy.
He pulled me out of my depression.
That experience pushed him over the edge.
The trial thrust O.J.'s attorneys into the limelight.
They handed me the job.
The democrats are trying to derail the republicans' legislative agenda.
The wind blew the door open.
Meteorites have dug out huge craters in the moon.
The alchemist wanted to turn lead into gold.
His political views were shaped by the depression.
The earthquake held up the project.
A rise in pressure accompanies a rise in temperature.
Smoking leads to cancer.
Cancer has been linked to smoking.
Russia replaced one government with another.
He carried the project to completion by himself.
They closed the door on a settlement.
He died from pneumonia.
Pressure goes up with temperature.
He ingratiated himself into the community.
Necessity is the mother of invention.
He went blind after his optic nerve was severed in the accident.
A settlement emerged from long discussions.
Difficulties began to arise.
The data **forced** me to change my theory. 
Trees in a forest grow toward the sun **in order to** get the light they need. 
They **made** a mountain **out of** a molchill. 
They are trying **produce** a new theory of physics. 
You reap what you **sow**.

John **killed** Bill.  
John **caused** Bill to die. 
John **caused** Bill pain. 
John **caused** trouble. 
John **had** Bill killed. 
John **brought it about that** Bill died. 
Her husband typed her thesis **because** he loves her. 
Her husband must love her – **because** he typed her thesis. 
The house caught fire **because** it was made of wood. 
Billy broke the window and ran away **because** boys will be boys.

All of these sentences express causation – but the not the same concept of causation. They do not all show the same inference patterns. As we shall see shortly, there are nearly two dozen ways to conceptualize causation, each with its own logic and vocabulary – and most of them metaphoric. Here is an all-too-brief tour of our mental resources for conceptualizing causation.

The Causes Are Forces metaphor can be seen most readily in the metaphorical view of Causation As Forced Motion, where States Are (conceptualized as) Locations (that is, bounded regions of space) and Change As Motion. For example, too many ‘pressures’ may ‘drive you into a depression’, and if you are lucky, you may ‘pull yourself out of the depression’. Here the cause ‘pressures’ are metaphorical forces, ‘drive’ and ‘pull’ are verbs of forced motion, applied metaphorically: linguistic expressions indicating forced-motion-to-locations can also be used to express caused-changes-of-states.

Consider the verbs of forced motion, such as ‘bring’, ‘take’, ‘push’, ‘pull’, ‘propel’, ‘throw’, ‘send’, ‘carry’, ‘drive’, and so on. Via the metaphors, Causes Are Forces, States Are Location, and Change Is Motion, they can all be used to express causation, conceptualized as caused-change-to-a-new-state. But each verb has a different logic. When you *bring* something, it accompanies you and you are applying force and control the whole way. But when you *throw* something, you apply force to an object initially and then it moves on its own. These literal logics of force are used metaphorically in causal sentences like ‘FDR *brought* the USA out
of the depression’. Here ‘FDR exerted force and control over the whole period, and the verb ‘throw’ cannot be substituted. Compare this with ‘The home run threw the crowd into a frenzy’. Here the home run initiated the frenzy, which then went on by itself. You cannot substitute the word bring. But you can say ‘The home run brought the crowd to its feet’, if the crowd rose to its feet while the home run ball was in the air.

The point here is that the source domain inferential structures concerning literal force are preserved in the target domain of general causation. There are kinds of causation with different inference structures, which are preserved under the metaphors.

The Causes Are Forces metaphor can combine with other metaphors as well. Take the metaphors Properties Are Possessions (as in ‘I have a headache’) and Change Is Acquisition or Loss (as in ‘I got a headache’ and ‘My headache went away’). The concept of physical giving involves applying force to an object so that it is transferred to another person, while taking involves applying force to an object so that it is transferred to you. Metaphorical giving and taking occur when Causes Are Forces combines with Properties Are Possession and Change Is Acquisition or Loss. Examples are ‘The noise gave me a headache’ and ‘The aspirin took my headache away’.

Notice how different these two conceptualizations of causation are. In caused metaphorical motion, the object of change is the patient that the force is applied to and that moves, as in ‘He put me into a dangerous position’. In caused metaphorical transfer, the object of change is the recipient of the giving, as in ‘He gave a lot of pain to his friends’. And just as the conceptualization of causation is different, so the grammar is different. In both cases, we have caused metaphorical motion, and so the grammar of forced motion can be used: Verb+direct object+direction prepositional phrase. But only with caused transfer can the double object construction indicating a recipient be used: ‘He gives his friends a lot of pain’.

Incidentally, the verb ‘cause’ is conceptually ambiguous. It can express causation as forced motion (as in ‘I caused him to leave’) or it can express causation as forced transfer (as in ‘He caused great pain to all his friends’ or ‘He caused his friends pain’). Depending on which concept of causation it is expressing, ‘cause’ takes the appropriate grammatical construction.

This is only the tip of the iceberg. There well over a dozen other forms of causation, with different inferences, different lexical items, and different grammars. In Causation As Forced Motion,
there is a path from an initial location to a new location, with a force moving you along the path. In the related Causal Path metaphor, just being on the appropriate path moving under your own steam will lead you to a result, with the path seen as the cause. Examples are ‘You’re on the road to ruin’, ‘If you keep going the way you’re going, you’ll become an addict’, and ‘Smoking leads to cancer’. Here prevention is getting someone or something off a causal path. Here the verb derail is appropriate, as in ‘They’re moving toward victory and will win unless we derail their campaign or they get bogged down’. A variant on the Causal Path metaphor is the Causal Source metaphor, as in ‘He died from pneumonia’. Here there is an implied Causal Path from an initial location to a resulting location.

Causation can also be conceptualized in terms of accompaniment, as in ‘Pressure increases with temperature’ or ‘An increase in pressure accompanies an increase in temperature’. This is the accompaniment that we find in the Forced Motion metaphor with the concept of bringing, and so it can be expressed with the verb ‘bring’, as in ‘An increase in temperature brought an increase in pressure’. Of course, verbs like ‘throw’ or ‘propel’ cannot be used here. One mode of accompaniment is being tied or linked together. And so we get, ‘A rise in cancer has been linked to mercury in fish, which has been linked to the mercury released into the air by coal-burning electric plants’. The concept of ‘Linking’ is what we use for conceptualizing ‘Causal Chains’.

When causation results in a change of form, we conceptualize it via Causes Are Forces, where States Are Shapes. Examples include ‘reshaping the bureaucracy’, ‘reforming politics’, and so on. When causation results in a new entity, we use creation concepts, such as ‘forging a new alliance’, or progerneration concepts, as in ‘The Internet gave birth to a new era of commerce’.

In the theory of essences, essences are taken to be the causal source of natural behaviour. There is a folk version of this philosophical theory, which shows up in sentences like ‘He broke the window and ran away, because boys will be boys’, or ‘The house caught fire because it was made of wood’.

Here is how all these apparently disparate metaphors hang together. The primary metaphor Causes Are Forces is at the centre of the system. It applies to cases of forced motion, forced transfer, and forced change of shape. In the bringing cases of Forced Motion, the causal force correlates with accompaniment along a causal path from a source to a resulting location. These overlaps provide conceptual links from Causes Are Forces to
Causal Paths, Causal Sources, Causal Links, and Causal Accompaniment, which is a version of Causes Are Correlations. The Causal Theory of Essences is an extension of Causes are Sources, a natural state (an essence) is a location on a Causal Path, where Nature is itself a force – an overwhelming force.

Finally, there are literal cases of various kinds, in which causation is characterized as having various degrees of directness of causal connection. In ‘John killed Bill’, the cause is direct, with cause and result as part of one event. In ‘John caused Bill to die’, the cause and the result are conceptualized as different events. In ‘John had Bill killed’, John is the ultimate cause, but there is an intermediate cause who performed the murder. In ‘John brought it about that Bill died,’ the cause is very indirect and not only a separate, but a distant event. This is expressed iconically in the grammar of English: The more direct the cause, the closer the result is to the cause in the sentence.

I will stop here. Chapter 11 of Philosophy in the Flesh has a great deal more detail.

There are several morals to these examples.

• We do not have a single concept of causation, but many, each with a different inference structure.
• The wide range of causal concepts and their inferences arises via metaphor.
• The central metaphor is Causes Are Forces. It combines with other metaphors to yield metaphorically complex causation concepts.
• Causation is embodied, via the exertion of force in everyday experience.
• There is no disembodied concept of causation with a single collection of inferences out there in the world.
• The world has many different patterns of occurrences. We call a large class of them instances of ‘causation’ because of the relationships we see with the way forces work.

What makes all these different concepts with different logics all cases of one thing? There appear to be two necessary conditions that follow from the Causes Are Forces metaphor. These conditions are:

1. The result cannot precede the cause. Under Causes Are Forces, this follows from the fact that forced motion cannot precede the application of the force.
2 The result would not have occurred without the cause. Under Causes Are Forces, this follows from the fact that forced motion does not occur without a force.

These two conditions define what Johnson and I call ‘a determining factor for a situation’. This is a necessary condition for causation: everything we conceptualize as a cause is a ‘determining factor for a situation’ in this sense.

But though this is a necessary condition, it is not sufficient. We can see this in cases like the following, which are not cases of causation.

His birth caused his death.
His marriage caused his divorce.
His taking the course caused his failure in it.

In each of these cases, there is a necessary condition that is not a cause. Birth is necessary for death but doesn’t cause it, marriage is necessary for divorce but doesn’t cause it. Thus, counter to the analyses of Lauri Kartunnen in terms of inference patterns and Richard Stalnaker (1968) and David Lewis (1973), in terms of model theory, these conditions do not define an abstract logical notion of causation, but only characterize necessary conditions. However, they are necessary conditions for all the concepts, including metaphorical concepts, that we understand as causal. It is these necessary but not sufficient conditions that conceptually tie together all the very different concepts that we understand as ‘causal’.

In my earlier career in generative semantics and logic, I introduced a fateful notation: CAUSE in capital letters. I assumed that there was a single abstract general concept of causation with a single logic. I also assumed that causation as a single unified phenomenon, with a single inference pattern, existed in the world and fitted the causal expressions in human language.

I was mistaken. The world contains lots of patterns. There is no single one that corresponds to all the situations we conceptualize as causal. We, because of the embodied metaphorical character of our minds, create the unified concept of causation. It is not just out there in the world.

This is a stunning discovery: it destroys the metaphysical realism that we were brought up to cherish – the idea of causation as a single objective structure or pattern in the world. From now on, we must always ask which of the two dozen or so metaphorical
concepts of causation is being used in a theory and which detailed logic of causation is inherited from embodied experience via those metaphors.

Anytime you hear the simple word 'causation' you should ask, 'Which one?' Anytime you hear a philosopher or psychologist claim to have found the one true causation, you should confront her with these examples.

Mathematics is Embodied and Metaphorical

The disembodied view of mind in Western thought has led to the idea of mathematics as disembodied. Indeed, a mythology has arisen around mathematics, a mythology that Núñez and I call The Romance of Mathematics:

The Romance
1 Mathematics is abstract and disembodied – yet it is real.
2 Mathematics has an objective existence, providing structure to this universe and any possible universe independent of, and transcending, the existence of human beings or any beings at all.
3 Human mathematics is just a subpart of abstract, transcendent mathematics.
4 Hence, mathematical proof allows us to discover transcendent truths of the universe.
5 Mathematics is part of the physical universe and provides rational structure to it. There are Fibonacci series in flowers, logarithmic spirals in snails, fractals in mountain ranges, parabolas in home runs, and π in the spherical shape of stars and planets and even bubbles.
6 Mathematics even characterizes logic, and hence structures reason itself – any form of reason by any possible being.
7 To learn mathematics is therefore to learn the language of nature, a mode of thought that would have to be shared by any highly intelligent beings anywhere in the universe.
8 Because mathematics is disembodied and reason is a form of mathematical logic, reason itself is disembodied. Hence, machines can, in principle, think.

Núñez and I, in Where Mathematics Comes From (2000) have argued in great detail that this Romance is false in every way. Instead, mathematics, like other products of the human mind, is embodied and mostly metaphorical. Along the way, we provide a
cognitive theory of mathematic ideas for advanced mathematics, and show how so-called 'abstract' ideas from higher mathematics are ultimately grounded in bodily experience and extended via conceptual metaphor.

Our argument begins with baby arithmetic, the small but real arithmetic capacities of newborns:

- the ability to subitize up to 4 objects – to distinguish 1 from 2 from 3 from 4 objects;
- the ability to do implicit baby arithmetic up to 4 – the ability to tell that if one object is present and you add one object, there should be two objects, not three, and so on.

This is determined in experiments that measure sucking rates and staring by babies in what is called the 'violation-of-expectation paradigm'. There are normal sucking rates and normal scanning patterns. Suppose a baby is presented with a stage with one puppet on it. The curtain then comes down and the mother shows the baby a second puppet and puts it behind the screen. The curtain goes up. If there are two puppets, the sucking rate and scanning pattern remain normal. But if there are three puppets, the sucking rate goes up noticeably and the baby stares at the stage (Wynn, 1992).

In short, such limited capacities seem to be part of the brain's neural circuitry at birth. The question we asked was how such limited baby arithmetic expanded into full-blown arithmetical concepts including laws of arithmetic. The mechanism, we argue, is metaphor. There are four basic metaphors for arithmetic that arise through normal everyday functioning:

1. The Object Collection metaphor, in which addition is adding objects to a collection and subtraction is taking them away.
2. The Object Construction metaphor, in which addition is putting objects together to form larger objects.
3. The Measuring Stick metaphor, in which sticks of the same size are laid out in a line to measure length.
4. The Motion Metaphor, in which addition is taking steps in a given direction, and subtraction is taking steps in the opposite direction.

These are primary metaphors that arise via conflation. For example, as a child puts objects in a container, she performs baby arithmetic on them and subitizes them. The same for, say, taking
steps in a direction. The conflation of these activities results in neural co-activation, which in turn results in the recruitment of neural connections. These neural connections are the physical mechanism by which metaphorical mapping is accomplished. When the metaphorical mappings are spelled out in detail, the Laws of Arithmetic turn out to be entailments of the mappings. For example, the law
\[ a + b = b + a \]
follows from the physical fact that you get the same number of objects in a container no matter what order you put them in.

Later, symbolic mappings are learned – mappings that link concepts to symbols and rules of symbolic computation to metaphorical inferences. Since conceptual metaphors preserve inference, the algorithmic rules of computation will always work.

What comes out of all this is a cognitive theory of arithmetic: how it extends from baby arithmetic to arithmetic laws; how the laws arise; what symbolization is; and why algorithms work. Included is a cognitive account of why

\[ -1 \times -1 = +1. \]

We go on to classes, sets, and logic. We show how the idea of a class is grounded in container image schemas and how the logic of classes is the logic of container schemas under the metaphor that A Class Is A Container-Schema. We then describe Boole’s metaphor for the algebra of classes in detail, in which he introduced the metaphorical mapping that creates the empty set from zero and the universal set from one. And we show how prepositional logic arises from a simple basic metaphor, namely, Propositions Are Classes Of World-States.

The point of all this is to demonstrate that formal logic is a cognitive construction and to show the detailed cognitive mechanisms by which logic is put together. Along the way we show that some of the stranger parts of set theory and logic arise from Boole’s metaphor. For example, via Boole’s metaphor which maps zero into the empty set, the arithmetic truth

\[ 0 \times n = 0 \]

maps onto the metaphorical statement ‘the empty class is a subclass of every class’, and that statement in turn maps onto the
strange statement ‘every proposition follows from a contradiction’. Thus, ‘the empty class is a subclass of every class’ and ‘every proposition follows from a contradiction’ are not part of the logic of the universe, but rather are strange metaphorical inferences following from Boole’s metaphor.

The heart of the book is our study of infinity. There we take up the question, ‘How is it possible for a human being with a finite mind to comprehend infinity?’ The answer again comes via metaphor.

There are two kinds of infinity – ‘potential infinity’ and ‘actual infinity’. Potential infinity is open-ended. You just start counting and keep on counting indefinitely. You start making a line segment longer, and you go on making it longer and longer indefinitely. You compute a number, say pi, to n decimal places, and then n+1, and then on and on indefinitely.

Actual infinity is a different matter. You don’t just keep counting and counting. You achieve the collection of all natural numbers, an infinity of them. You don’t just keep extending the line longer and longer. You reach the point at infinity where all parallel lines meet in projective geometry. You don’t just keep getting more and more decimal places for pi without reaching pi. Instead, you reach pi by getting the infinity of decimal places, all of them. This is actual infinity. This is the concept of infinity that is interesting in mathematics and on which a huge amount of mathematics rests.

The question is, how can a finite being conceptualize and reason about actual infinity in all its forms – infinite sets, points at infinity, infinite decimals, transfinite numbers, infinitesimal numbers, limits, least upper bounds, and so on?

The answer begins with aspect in linguistics. As we saw in Chapter 3, Sriniv Narayanan’s computational neural model of motor control characterizes the conceptual structure of aspect. Cases of potential infinity fit what, in linguistics, is called ‘iterative imperfective aspect’: There is an initial state, then perhaps a starting action, then a repeated central action that goes on and on indefinitely, without end. An example would be counting by ones without stopping: 1, 2, 3, … and forming sets of numbers as you go along: {1}, {1, 2}, {1, 2, 3}, …

Compare this with an ‘iterative completive aspect’: There is an initial state, then perhaps a starting action, then a repeated central action that goes on for a finite number of iterations, then stops, resulting in a final state. An example would be counting by
ones forming sets as you go along and stopping at ten, with the set of the first ten integers.

Let us now construct the following metaphor:

• The target domain is an iterative imperfective action: It has an initial state, a starting action, and a central action that iterates.
• The source domain is the corresponding iterative completive action: It has the same initial state, the same starting action, the same central action that iterates, and a final state (reached after some finite number of iterations).
• The metaphor maps the initial state to the initial state, the starting action to the starting action, and the iterated central action to the iterated central action.
• In addition, the metaphor maps the source domain final state onto a corresponding final state in the target domain, creating a target domain final state via the metaphor.

This last step, which adds a metaphorical final state to cases of potential infinity, produces actual infinity – infinity as a thing, an entity. Moreover, this mapping also imposes certain source domain inferences on the target domain. In the source domain, the final state comes after all the earlier states, and it is the first state that comes after all the earlier states. The metaphorical mapping imposes this inferential structure on the target domain, with the result that the final state of actual infinity comes after all earlier states and is the first state to come after all earlier states. These inferences characterize crucial properties of actual infinity.

Núñez and I call this metaphor The Basic Metaphor of Infinity. It is fully general, and need not be about mathematics at all. It has an open-ended number of special cases, in which the initial state, the starting action, and the iterated central action can be spelled out in detail.

Note that the concept of actual infinity is not literal. The target domain has no literal connection between the earlier cases that go on without end and the final state of actual infinity. The connection between the open-ended cases of potential infinity and the final state of actual infinity is not made in the target domain. The connection is made in the source domain and projected via the metaphor.

Núñez and I show how this metaphor, when fleshed out with the right special cases, characterizes a wide range of the cases of infinity in various subject matters. For example, here's how we would characterize the set of all natural numbers:
Let the initial state be the empty set. The starting action is to produce the number 1 and form a union of the set containing 1 with the empty set. The iterated central action is add 1 to the previously produced number and form that union of the set containing that number with the previously formed sets. The final state is the set of all natural numbers.

In this special case, the fact that the final state comes after all the other states has as a special case the fact that all the previously formed sets are subsets of the set at the final state, and this is the smallest such set. This is the infinite set of all the natural numbers.

Now take the example of the point at infinity in projective geometry:

Let the initial state be an isosceles triangle, ABC, with AC = BC. Let the starting action be to double the length of AC and BC. Let the iterative action be to take the previous length of AC and BC and double it.

After each iteration, the angles CAB and CBA are closer to 90° than after the previous iteration.

At the final state, we still have an isosceles triangle, ABC, with AC = BC. AC and BC are infinitely long. C is therefore a point 'at infinity'. Angles CAB and CBA are 90°. AC and BC are thus parallel lines that meet at infinity, point C. Since the distance AB and the orientation of the triangle were arbitrarily chosen, the initial state of this instance of the metaphor fits all isosceles triangles, and in the final state, all parallel lines meet at infinity.

In our book, we go through many other cases: inversive geometry, mathematical induction, limits, infinite sums, least upper bounds, transfinite cardinals and ordinals, and infinities. The Basic Metaphor of Infinity is precisely stated, as are each of the special cases. The result is a precise cognitive account of actual infinity for a wide range of cases in mathematics. Among those cases are four different kinds of infinite numbers.

There are two morals here:

1 Actual infinity is an embodied metaphorical human concept. It has to do with human motor control and the human concept of aspect, as well as conceptual metaphor.
2 Actual infinity is not a ‘thing’ in the world, nor are infinite numbers, nor are other mathematical entities.

In the last four chapters of our book we work out the metaphorical structure of Euler’s classic equation:

\[ e^\pi + 1 = 0. \]

We show what this equation means and why its truth follows from what it means, in full detail, and in the process we set out a theory of mathematical ideas. What we learn from this is that mathematical concepts are embodied, just like other concepts. So-called ‘abstract’ mathematics arises by conceptual metaphor and/or conceptual blends – the basic cognitive mechanisms underlying abstract concepts in general.

What we show in detail throughout the book is that the Romance of Mathematics that we started with is false in every detail. Mathematics is not objectively true. Mathematics is not out there in the world structuring the universe. There are no parabolas out there in Barry Bonds’ home runs, no logarithmic spirals in snails, no Fibonacci series in flowers. Mathematical logic is not out there providing a rational structure to the universe. Mathematics – including mathematical logic – comes out of us.

Why can mathematics work in the sciences? Not because it is out there structuring the laws of nature. Instead, mathematics, via its metaphors and conceptual blends, expresses ideas, many of them ordinary everyday concepts – concepts such as change, proportion, size, inverse, change in inverse proportion to size, and so on. There is a mathematics of change, a mathematics of proportion, a mathematics of change in inverse proportion to size.

Scientists are careful observers of the world and categorize phenomena in terms of ordinary everyday concepts such as change, proportion, size, inverse, change in inverse proportion to size, and so on. Careful scientists use their concepts to categorize physical phenomena and they use the mathematics appropriate to those concepts to do their calculations.

There is nothing mysterious here. The fit between mathematics and the physical world comes in the minds of scientists.
Morality

Kant was wrong about morality – very wrong. Morality cannot follow from universal transcendent reason, because there is no universal transcendent reason. Reason is not transcendent; it comes out of our embodied experience.

To understand what is meant by morality, we must begin empirically. How do people conceptualize morality? What are the conceptual metaphors through which morality is understood? What is the cognitive structure of a moral system? These are the empirical questions. Moral questions arise only after these are answered.

The conceptual structure of morality, so far as Mark Johnson and I have been able to discern, comes in the following parts:

• A collection of primary metaphors for morality that are widespread around the world.
• Moral systems, based metaphorically on models of the family and of childrearing, that organize the primary metaphors and assign them priorities.
• Interactions with other commonplace metaphors.

Let us begin with the primary metaphors for morality.

Morality is fundamentally about well-being and human flourishing. There are certain common correlations between well-being and ordinary experiences that give rise to primary metaphors for morality. Those correlations and the primary metaphors they give rise to are as follows:

• It is better to be strong than to be weak.  
• Morality Is Strength; Immorality Is Weakness.
• It is better to be healthy than to be sick.  
• Morality is Health; Immorality Is Disease.
• It is better to have possessions you need than not to have them.  
• Well-being Is Wealth.
• It is better if debts are paid than if they aren’t.  
• Morality Is Accounting (Balancing The Moral Books).
• It is better to be able to stand upright than not to be able to.  
• Moral is Up; Immoral Is Down.
• It is better to function in the light than in the dark.  
• Morality Is Light; Immorality Is Darkness.
• It is better to be clean than to be dirty.
• Morality Is Cleanliness; Immorality Is Filth.

• It is better to have pure food than rotten food.
• Morality is Purity; Immorality Is Rottenness.

• It is better to be cared about than not cared about.
• Morality Is Empathy; Immorality Is Not Caring.

• Beauty is better than Ugliness.
• Morality Is Beauty; Immorality Is Ugliness.

• It is better to be whole than not to be.
• Morality Is Wholeness; Immorality Is Degeneration.

• It is better to be nurtured than to be neglected.
• Morality Is Nurturance; Immorality Is Neglect.

• It is better to be accepted by a community than not to be.
• Morality Is Community Acceptance.

• Fair treatment is better than unfair treatment.
• Morality Is Fairness.

• It is better to obey your parents than not to.
• Morality Is Obedience; Immorality Is Disobedience.

• It is better to stay where people are than to stray.
• Morality Is Staying within Boundaries; Immorality Is Transgressing.

• Morality Is Staying on a Path; Immorality Is Straying From The Path.

• It is better if moral people have power than if immoral people do.
• The Moral Order: Morality Is Preserving Traditional Power Relations.

• It is better for people to be good by nature than not to be.
• Moral Essence: Morality Is Character (or Virtue).

For the most part, these should be obvious. A few need further explication.

Well-being Is Wealth combines naturally with Morality Is Accounting. If you do something good for me, then I am ‘in your debt’ and may ask ‘How can I ever repay you?’ or say ‘I owe you one’. If you do something bad to me, then you owe me, since by metaphorical moral arithmetic, giving something bad is the equivalent to taking something good. In this case, the moral
books can be balanced in various ways. (1) Restitution: You do something equally good to make up for it. (2) Retribution: I do something bad to you. (3) Revenge: I take something good from you. (4) Forgiveness: I cancel the debt. Our notion of justice is a form of moral accounting.

Moral Essence is the basis of virtue ethics: Those who are good by nature or are raised to be virtuous, will have ‘character’ and will act morally.

The Moral Order is a conservative metaphor: It says that traditional hierarchies of power were there for good reason and should be maintained. American conservatives have a moral hierarchy that includes: God over Man; Man over Nature; Adults over Children; America over Other Countries; Western Culture over Other Cultures – and in more extreme cases the hierarchy extends to: The Wealthy over the Poor; Men over Women; Whites over Non-whites; Straights over Gays; and Christians over Non-Christians.

These primary metaphors for morality are widespread around the world. One can find versions of them in many cultures. But each is isolated. Among the things that allow them to fit together into a moral system are, interestingly enough, models of the family. In Moral Politics (1996/2002), I discovered that major political ideologies in America (and perhaps elsewhere) are based on ideal models of family life, and that such family-based models commonly give structure to moral systems by giving priorities to primary metaphors for morality.

The major family models I have investigated to date are the Strict Father model, which is the basis of conservative morality and politics, and the Nurturant Parent model, which is the basis of progressive morality and politics.

The Strict Father Model

Assumptions: The world is dangerous and children are naturally bad. The father’s job is to protect and support the family. He is the moral authority. He teaches his children right from wrong, which are taken to be absolute. The child’s job is to obey. Punishment is required to balance the moral books. The father’s moral duty is to physically punish his children when they do wrong, assuming physical discipline in childhood will develop the internal discipline adults need to be moral people and to succeed. Children are to become self-reliant through discipline and the pursuit of self-interest. When children are
mature, they are on their own and parents are not to meddle in their lives.

In this model ‘character’ means moral strength. Moral weakness or dependency is seen as a sign of lack of discipline and the inability to lead a moral, self-reliant life. The Moral Order makes sense here, since worldly power is a reflection of the discipline needed to lead a moral life.

This family model can be applied metaphorically so that other figures can become the moral authorities. In a common model, it is the Church that becomes the moral authority. Kantian morality is based on the metaphor of Universal Reason As The Strict Father who tells you what is right and wrong and whom you are to obey. Kant thus places moral authority in each of us, while keeping an absolute right and wrong.

The major metaphors for morality that tend to accrue to this model are Morality As Strength, Obedience, Staying Within Boundaries, Retribution, Purity, Cleanliness, Uprightness, Health, Wholeness, and the Moral Order.

The Nurturant Parent Model

Assumptions: Both parents are equally responsible for running the household and raising the children. Their job is to nurture their children and raise their children to be nurturers. To be a nurturer you have to be empathetic and responsible (for oneself and others). Responsibility implies competence, education, hard work, protection, and social connectedness. Empathy requires (1) fairness, (2) open, two-way communication, (3) a happy, fulfilled life (unhappy people are less likely to want others to be happy), and (4) restitution rather than retribution to balance the moral books. To promote happiness and fulfilment, aesthetics is important. In the place of specific strict rules, there is a general ethics of care that says: Help, Don’t Harm. To be of good character is to be empathetic and responsible.

The major metaphors for morality that tend to accrue to this model are Morality As Empathy, Nurturance, Connection To Others, Fairness, Light, and Beauty.

Strictness and Nurturance are two fundamentally different approaches to morality. Different people may even take them as defining what morality is. Many of the most basic moral issues in Western society come down to issues of strictness versus nurturance. The gap between these approaches appears to be irreconcilable.
Strict Morality comes with four requirements on the human mind and language:

1. **Strict categorization**: Everything is either in or out of a category.
2. **Literalness**: Moral rules, and criteria for determining whether they are followed, must be literal.
3. **Perfect communication**: Moral rules are assumed to be clearly understandable.
4. **Folk behaviourism**: People normally act so as to get rewards and avoid punishment.

We know from cognitive science (especially cognitive linguistics) that all four of these are false. The mind and language just don't work that way. Strict morality is simply incompatible with what we have learned about the embodied mind.

The study of the embodied mind and metaphorical thought have a great deal to tell us about morality and ethics. First, they allow us to gain insight into what morality and ethics are taken to mean. Second, they place cognitive constraints on moral action. A view of what human beings should do that is fundamentally at odds with the nature of the human mind and language is so out of touch with reality that it must be rejected. Nurturant morality, on the other hand, is consistent with what is known about the mind and language. It is at least a candidate for a viable approach to moral action.

**Religion**

All religion has a moral dimension. The major religions in the West – Judaism, Islam, and Christianity – have both strict and nurturant versions. My 1996 book *Moral Politics* contains a description of both versions of Christianity. The Torah and Talmud contain both approaches, though most American Jews today tend towards the nurturant approach. Fundamentalists in all religions, of course, take the strict approach; an ideological commitment to strict morality is what makes them fundamentalists. Religious disputes in the West often centre on strictness versus nurturance. To understand any approach to religion, you need to understand the conceptual structure of its approach to morality. Cognitive science can be quite useful here.
Does God Exist?

The following story is told of the great intellectual historian, A. O. Lovejoy. He was being considered for the presidency of Johns Hopkins University. At his interview, a member of the Board of Trustees asked him if he believed in God. His response was, ‘Which one?’ and he reeled off a hundred conceptions of God.

It does seem sensible to understand a question before trying to answer it. Since God is presumably ineffable, God can only be approached through metaphor. All understandings of God make use of metaphor.

I have been looking at metaphorical conceptions of God, and so far I have found three major types.

1. God As Parent, or as having parent-like qualities: Eve Sweetser, in her study of the Yom Kippur liturgy, found that the conceptions of God explicitly given there formed a radial category with God As Father in the centre. The properties of a parent are (a) a powerful authority figure; (b) someone who loves you; (c) someone to whom you owe your existence; (d) a protector and guide; (e) someone who prescribes and judges your behaviour; (f) the source of life; and (g) someone who has chosen to bring you into the world and take care of you – a chooser. The metaphors are: God As (a) King, Lord; (b) Lover; (c) Creator, Potter, Jeweller; (d) Shepherd; (e) Lawgiver, Judge; (f) Breath (the source of life); and (g) the Chooser of the Chosen People.

2. God As The Infinite: The Basic Metaphor of Infinity is not just about mathematics. It has many special cases. It is commonly used to conceptualize God as infinite. Special cases include: God as (a) All-Seeing; (b) All-Knowing; (c) All-Powerful; (d) All-Good; and (e) the First Cause.

3. God As Immanent: The Kabbalah says: ‘Do not say, “This is a stone and not God.” God forbid! Rather all existence is God and the stone is a thing pervaded by divinity.’ The idea that God is all of existence is also central to the traditions of pantheism and naturalism in Christianity. It has the consequence that empathic projection onto anything and anyone is contact with God.

My friend, Zoketsu Norman Fischer, former Abbot of the San Francisco Zen Center, has pointed out to me that there is a generalization across these three metaphors. All of them are attempts
to connect with what is outside and beyond you. His point is that that is what religion and spiritual life in general are about.

Suppose, like Lovejoy, I were asked, ‘Do you believe in God?’ Knowing that one can only conceptualize God via metaphor, I would go through the list, metaphor by metaphor. ‘God as Father, King, Jeweller, Shepherd…’ I can understand the appeal of the personification metaphors, but I can’t bring myself to believe those metaphors as if they were literally true. Moreover, it’s hard to believe that anyone really takes them literally, really believes that God Is Daddy! It is hard to believe that anyone takes ‘The Lord Is My Shepherd’ as literally true; after all, that would make you a sheep—literally, with wool and eating grass. It is taken as a metaphor.

Next, God As The Infinite. Understanding the Basic Metaphor of Infinity, I fully appreciate the beauty and preciousness of actual infinity as an idea. But I cannot take actual infinity as an objectively existing thing in the world.

Finally, God As What Exists—the world as it is. Well, what exists exists. Metaphor number 3 has a lot going for it. It makes the world sacred. I do believe the world is sacred. Metaphor 3 comes with an appropriate morality: nurturant morality, which is structured around empathy and responsibility. Empathy is a neural matter, arising from the ability of mirror neurons to allow you feel what others feel, to connect you to others and to the world.

Do I have a soul separate from my body? Given that it couldn’t perceive, think, be conscious, empathize, or have my personality or character, I wouldn’t want to call it a ‘soul’ even if it did exist.

What about spiritual experience? During meditation practice, I have what I would call spiritual experience. But, as James H. Austin argues in Zen and the Brain (1999), spiritual experience is a form of physical experience. There is nothing unscientific about it. It is quite bodily. And it has to be that way, because it is only through your neural system that you can experience anything at all.

Is there a metaphor for God that I— as a scientist, a cognitive scientist— can accept? I used to think that the answer was ‘No’, since I was thinking of God only as given in Metaphors 1 and 2. I thought I was an atheist, but now I realize that I am not, since I can say ‘Yes’ to Metaphor 3: God As Immanent, as what exists.

What is sacred is the world. It has inherent and ultimate value. I take spiritual connection as empathy with the world (via mirror and canonical neurons), empathy with people, and with what David Abram calls ‘the more than human world’: plants, animals, soil, sea, clouds, geological formations, all of it, including the wonderful inventions of human minds. With it comes a way of life
– a commitment to nurturant morality, that is, a commitment to empathy and responsibility. That includes a commitment to a practice, a quite demanding practice of rigorous thought, finding out about and trying to comprehend the world, meditation, exploration of self and others, appreciation, connection, giving, and experiencing fully the wonders of your own mind and body and the wonders of the Other – all kinds of Other.

The study of the embodied mind and metaphorical thought is a liberating enterprise. It frees us from the many tyrannies of the disembodied literal mind:

• the tyranny of a literalist philosophy that devalues the imaginative powers of human minds and mistakenly sees only one causation;
• the tyranny of a falsely disembodied mathematics that would turn the beauty of human creation into mere discovery;
• the tyranny of a literalist science that ignores the contribution of human bodies and minds and culture in constituting knowledge;
• the tyranny of a strict morality centred on authority rather than empathy and responsibility;
• the tyranny of a disembodied spiritual life whose focus is outside the magnificence of this world; and
• the tyranny of an unreal disembodied ‘realism’ that removes from us the proper responsibility for the reality we bring into being through living by our metaphors.

An embodied realism is liberating. Recognizing metaphorical thought acknowledges what is real. Understanding how our bodies shape even mathematical and scientific thought makes human knowledge that much more stunning, worthy of awe, and worth pursuing.

Best of all, cognitive science and neuroscience explain what is unsatisfying (at least to me) in religions and spiritual traditions based around personification, infinity, and essence metaphors. At the same time, they clarify the rich spiritual and moral life of those who find what is sacred and of ultimate value in the here and now, in the connection to the Other – the human Other and the more than human Other.

Do not say God is not in the stone and the oak, the virus and the bacterium, the ant and the aardvark, you and me, houses and chairs, books and wine. Sex and eating are sacred, they unite you with the Other, hence with God under Metaphor 3. Travelling the world is sacred; it is an exploration of God under Metaphor 3. Preserving the world is sacred; it is preserving God under
Metaphor 3. Creative activity is sacred; it is creating God under Metaphor 3. All experience is sacred; it is the experience of God under Metaphor 3.

‘God’ has largely been defined by Metaphors 1 and 2. That is unfortunate, because it has located the sacred in a place other than in the world and in our relationships to the Other – human and more than human: animal, plant, soil, water, air. And what is not sacred is open to destruction in the name of what is.

We are our bodies. We don’t need disembodied souls to have full moral and spiritual lives. Human bodies are more than enough.

Conclusion

Research on the embodiment of mind is anything but esoteric and irrelevant. It has the deepest consequences not just for academic fields like Philosophy, Psychology, Linguistics, Mathematics, and Literary Studies, but also for how we understand our experience, govern our societies, live our lives, and so change our lives. The new understanding of morality, politics, and religion coming out of our understanding of mind is critical if we are to live better lives and leave the world better than we found it.

References


