1 Introduction: what makes science possible?

Peter Carruthers, Stephen Stich and Michael Siegal

In this brief opening chapter we briskly review some of the recent debates within philosophy and psychology which set the stage for the present collection of essays. We then introduce the essays themselves, stressing the inter-linking themes and cross-connections between them.

1 Introduction

The central position of science in our contemporary world needs no emphasis. Without science (broadly construed, to include all forms of technical innovation) we would still be roaming the savannahs of Africa like our *Homo habilis* ancestors, digging up tubers and scavenging scraps of meat. And without science (construed narrowly, as involving the application of an experimental method) we would have seen none of the advances in knowledge, technology and accumulation of wealth which have transformed the world and most of its people in just the last four centuries or so. Science now touches every aspect of our lives, from cradle (indeed, sometimes from conception) to grave. Given the manifest importance of science, the search for a scientific understanding of scientific thought and activity itself should need no further motivating. But in fact, the attempt to explain scientific cognition is not only extremely hard, but raises a whole host of fascinating and puzzling questions about the nature, development and operations of the human mind, and its interactions with culture.

This book is about the question: what makes science possible? Specifically, what features of the human mind, of human cognitive development and of human social arrangements permit and facilitate the conduct of science? These questions are inherently inter-disciplinary, requiring co-operation between philosophers, psychologists and others in the social and cognitive sciences. And they are, it should be stressed, questions which are as much about the psychological underpinnings of science as they are about science itself. That is, they concern what it is about our minds and/or mind-guided social interactions which make science possible, and how these factors relate to other things which we can do, either as adults or children. Indeed, one of the important themes of the book is the broad-scale architecture of the mind. For in order to understand

how science is possible we have to understand how our capacity for scientific theorizing fits into the structure of the mind, and what consequences that might have for the practice of science.

Steven Pinker in his well-known book *How the Mind Works* (1997) raises questions to which this volume is a partial answer. Having argued that cognitive science is approaching an understanding of many different aspects of the mind – vision, language, memory, and so on – he then lists various factors about human beings which (he says) we *cannot* yet begin to explain. One of these is consciousness. Another is science. According to Pinker, we don't really have any grip on how human beings can be capable of scientific thinking and reasoning. That is a striking and challenging claim. Taking up and answering that challenge, as this book begins to do, is something which should be of interest, not only to psychologists and philosophers of psychology, but to all those interested in understanding either the nature of the human mind, or the nature of science, or both.

2 Philosophy of science: a very short recent history

As the title of this section suggests, we here take the reader through a *very* brisk tour of recent developments in the philosophy of science.

2.1 Positivism and beyond

In the beginning of our story there was logical positivism, which dominated much of the middle part of the twentieth century (Ayer, 1946; Carnap, 1950, 1967; Hempel, 1965). The logical positivists were heirs to the classical empiricist tradition in philosophy of science and theory of knowledge – believing that all enquiry should be grounded in observation, and that the scientific task is essentially that of accommodating existing observations while correctly predicting new ones. They also believed that one central paradigm of scientific enquiry is enumerative induction. A pattern is discerned in our observations (each raven so far observed has been black), and then generalized into a universal law (all ravens are black). Much intellectual effort was expended in attempts to justify our inductive practices, and in discussion of the problem of underdetermination of theory by data (there are always, in principle, infinitely many distinct theories – that is, generalizations of the data – consistent with any finite data-set).

One stalwart and long-standing critic of logical positivism was Popper (1935, 1963, 1972). Popper pointed out that induction can only *generalize from* observation, whereas science characteristically goes *beyond* experience, by postulating theoretical entities such as electrons and X-rays, for example, which might explain it. The method of science, Popper argued, is not observation and

induction, but rather non-algorithmic theory-construction followed by testing. Scientists invent theories to *explain* their data, using imagination, analogy, intuition and any other resources which may happen to come to hand. (And since explanation isn't easy, scientists will generally be satisfied if they can construct just *one* explanatory theory.) But then, having devised a theory, they subject it to rigorous testing – deriving from it predictions concerning the observations which might be expected if the theory were true, and attempting to falsify the theory by making those observations which seem least likely. A theory is justified when it escapes falsification, on this view.

Up to this point, philosophy of science had been conducted in a relatively *a priori* fashion – with some reference to real scientific examples, admittedly, but mostly with philosophers of science just thinking about what scientists *ought* to do, rather than about what they actually *do* do. This all began to change in the 1960s and 1970s, when philosophy of science took its so-called 'historical turn', through the work of Kuhn (1962), Feyerabend (1970, 1975) and Lakatos (1970).

2.2 *Historical and naturalistic turns*

As Kuhn and others noticed, when one studies the history of science one discovers that the behaviour of actual scientists often fails to conform to the norms of scientific method laid down by philosophers of science. In particular, when scientists know of data inconsistent with their theories, they do not immediately abandon those theories and start again, as Popper would have had them do. Sometimes they try to explain away the recalcitrant data while continuing to hold onto their theory; but as often as not they simply ignore it, and get on with the business of developing their favoured theoretical approach. This gives rise to a dilemma for *a priori* philosophers of science. Either they can claim that the actual practice of scientists has been irrational, or at least inappropriate – in which case the immense success of science is rendered utterly mysterious. (How *can* science be so successful if scientists have mostly been doing it all wrong?) Or they can take the historical results as a refutation of their proposed methodologies. Almost all philosophers of science converged on the latter course – and wisely so, surely.

At about the same time as, and not unrelated to, the historical turn in philosophy of science, much of philosophy was undergoing a 'naturalistic turn'. This took place in epistemology and philosophy of mind generally, as well as in the philosophy of science in particular. Most philosophers started to accept as a serious constraint on their theorizing, that both human mental processes and human modes of acquiring knowledge are *natural*, happening in accordance with causal laws as do all other events in nature; and that philosophical attempts to achieve an understanding of the nature of these processes should be seen as continuous with scientific enquiry. This resulted in a plethora of

causal theories of mental and epistemic phenomena – with the development of causal theories of reference (Kripke, 1972; Putnam, 1975), of memory (Martin and Deutscher, 1966; Locke, 1971), of perception (Grice, 1961; Dretske, 1969; Goldman, 1976), of knowledge (Armstrong, 1973; Dretske, 1981; Goldman, 1986), and of justification (Goldman, 1979, 1986). Indeed, the dominant theory of the overall nature of the mind, which rose into ascendancy during this period, was functionalism, which sees mental states and events as individuated by their characteristic causal roles (Putnam, 1960, 1967; Lewis, 1966; Armstrong, 1968 – see section 3 below).

It became important, then, to see science, too, as a natural phenomenon, somehow recruiting a variety of natural processes and mechanisms – both cognitive and social – to achieve its results. Philosophers of science began to look, not just to history, but also to cognitive psychology in their search for an understanding of scientific activity (Nersessian, 1984b, 1992a; Giere, 1988, 1999a; Thagard, 1992). This trend is continued into the present volume, with a number of its philosophical authors appealing to psychological models or data in their chapters (e.g. Carruthers, chapter 4; Faucher *et al*., chapter 18; Giere, chapter15; Nersessian, chapter 7).

2.3 Science and the social

Our story so far has mostly been one of good news – with philosophy of science in the last century, like science itself, arguably progressing and/or getting somewhat closer to the truth. But one out-growth of the historical turn in philosophy of science was a form of social constructivism and relativism about science (Bloor, 1976; Rorty, 1979; Latour and Woolgar, 1986; Shapin, 1994). On this view, scientific theories are proposed, accepted and rejected in accordance with a variety of political and social forces which needn't have any connection with truth or reality, nor with reliable standards of evidence and rationality. Indeed, on this account the very idea of 'reality' as something set over and against our various socially constructed perspectival representations is unintelligible. The only real sense which can be made out of one theory being better or worse than another is in terms of its political–social influence.

While social constructivism has not found wide acceptance among philosophers of science generally (nor among the contributors to this volume in particular), it has perhaps played a useful role in emphasizing the social dimension of science and scientific activity. And one characteristic of recent work has been to combine elements from the social constructivist position with an overall realism about science – allowing that science progresses through varied social interactions among scientists, while arguing that those interactions can still be such as to facilitate increases in knowledge of the world (Kitcher, 1993, chapter 14 in this volume; Goldman, 1999; Thagard, 1999).

If one had to characterize the current state of play in contemporary philosophy of science – using very broad brush-strokes, of course – it would be that it is naturalistic and broadly realist in orientation; interested in both descriptive and explanatory accounts of scientific activity (historical, psychological and sociological); but also concerned with normative issues concerning the probity and/or reliability of various scientific practices. That picture pretty much represents, too, the orientation of most of the contributions to this volume.

3 Philosophy of mind: another short recent history

Since this book is as much about the mind as it is about science, in this section we provide a brief overview of the main developments to have taken place in the philosophy of mind over the last half-century or so.

3.1 Behaviourism and beyond

If our story of recent philosophy of science began with logical positivism, then our account of recent philosophy of mind has to begin with logical behaviourism, which also dominated much of the middle part of the twentieth century (Ryle, 1949; Wittgenstein, 1953). The leading idea of behaviourism is that it is a mistake to treat talk about the mental as talk about inner causes of overt behaviour. To think in this way, according to Ryle, is to commit a kind of *category-mistake*. Talk about the mental is not talk about mysterious inner causes of behaviour, but is rather just a way of talking about dispositions to behave and patterns of behaviour.

Behaviourism did have some attractions. It allowed humans to be included smoothly within the natural order by avoiding postulation of anything 'ghostly' inside the organic machinery of the body. It was thus able to reject any sort of ontological dualism, between non-physical minds and physical bodies. For the main objection to such a dualism has always been the problem of explaining how there can be any sort of causal commerce between the states and events of a non-physical mind and those of a physical brain or body.

The deficiencies of logical behaviourism were even more apparent, however. There were two main problems. One is that it seems quite implausible that knowledge of one's own mind could consist in knowledge of one's behavioural dispositions, since this hardly leaves any room for the idea of first-person authority about, or any kind of privileged access to, one's own thoughts and feelings. (Hence the old joke about the two behaviourists who meet in the street – 'You're feeling fine', says one, 'But how am I?')

The other major deficiency is this: logical behaviourism was offered as a piece of conceptual analysis. It was supposed to be an account of what had all along been the import of our psychological discourse. That being the Rylean

stance, a serious criticism of logical behaviourism is that it fails on its own terms, as an exercise in analysis. According to behaviourism what look like imputations of internal mental events or states should actually be construed as 'iffy' or conditional statements about people's actual and possible behaviour. The objection to the pretensions of behaviourist conceptual analysis, then, is that nobody has ever actually produced a single completed example of the behavioural content of such an analysis.

Indeed, there are principled reasons why no such behavioural analysis can be provided. For as Davidson (1970) pointed out, a particular belief or desire only issues in action together with, and under the influence of, other intentional states of the agent. There is no way, therefore, of saying what someone who holds a certain belief will do in a given situation, without also specifying what other beliefs and desires that agent holds. So analysis of a belief or a desire as a behavioural disposition requires invoking other beliefs and desires. This point has convinced practically everyone that Ryle was wrong. A belief or a desire does not just consist in a disposition to certain sorts of behaviour. On the contrary, our common-sense psychology construes these states as internal states of the agent which play a causal role – never singly, but always at least jointly – in *producing* behaviour.

3.2 Physicalism and functionalism

With dualism and logical behaviourism firmly rejected, attempts since the 1960s to give a philosophical account of the status of the mind have centred on some combination of physicalist *identity theory* with *functionalism* of one or another sort.

There are two distinct versions of identity theory which have been the focus of philosophical debate – *type-identity* theory and *token-identity* theory. According to the former, each type of mental state is identical with some type of brain state – for example, pain is the firing of C-fibres. According to token-identity theory, in contrast, each particular mental state or event (a 'token' being a datable particular rather than a type) is identical with some brain state or event, but it allows that individual instances of the same mental type may be instances of different physical types.

Type-identity theory was first advocated as a hypothesis about correlations between sensations and brain processes which would be discovered by neuroscience (Place, 1956; Smart, 1959; Armstrong, 1968). Its proponents claimed that the identity of mental states with brain states was supported by correlations which were just starting to be established by neuroscience, and that this constituted a scientific discovery akin to other type-identities, such as *heat is molecular motion*, *lightning is electrical discharge* and *water is* H_2O.

Most philosophers rapidly came to think that the early confidence in such type-correlations was misplaced, however. For consider a sensation type, such as pain. It might be that whenever *humans* feel pain, there is always a certain neurophysiological process going on (for example, C-fibres firing). But creatures of many different Earthly species can feel pain, and it is also possible that there are life-forms on other planets which feel pain, even though they are not closely similar in their physiology to any terrestrial species. So, quite likely, a given type of sensation is correlated with lots of different types of neurophysiological state. Much the same can be argued in the case of beliefs, desires and other mental kinds.

The conclusion drawn from these considerations was that type-identity theory is unsatisfactory, because it is founded on an assumption that there will be one-one correlations between mental state types and physical state types. Rather, we should expect mental state types to be *multiply-realized* in physical states. This is just what the thesis of token-identity affirms: each token mental state is identical with some token physical state; but instances of the same mental state type can be identical with instances of different physical types.

At about the same time, and connected with these debates concerning mind–brain identity, *analytic functionalism* was proposed as an account of the manner in which we conceive of mental states. The guiding idea behind functionalism is that some concepts classify things by what they *do*. So transmitters transmit, while aerials are objects positioned to receive air-borne signals; and wings are limbs for flying with, while eyes are light-sensitive organs for seeing with, and genes are biological structures which control development. Similarly, then, it was proposed that mental concepts are concepts of states or processes with a certain *function*, or distinctive causal role (Putnam, 1960, 1967; Lewis, 1966).

Functionalism seemed to be the answer to several philosophical prayers. It could account for the multiple realizability of mental states, since physiological states of a number of distinct types could nevertheless share the same causal role. And it had obvious advantages over behaviourism, since it accords much better with ordinary intuitions about causal relations – it allows mental states to interact and influence each other, rather than being directly tied to behavioural dispositions. Finally, it remains explicable that dualism should ever have seemed an option. For although (according to functionalists) we conceptualize mental states in terms of causal roles, it can be a contingent matter what actually *occupies* those causal roles; and it was a conceptual possibility that the role-occupiers might have turned out to be composed of some sort of *mind-stuff*.

There were two main problems with analytical functionalism, however. One is that it is committed to the analytic–synthetic distinction, which many philosophers think (after Quine, 1951) to be unviable. And it is certainly hard to decide

quite *which* truisms concerning the causal role of a mental state should count as analytic (true in virtue of meaning), rather than just obviously true. (Consider examples such as: that *belief* is the sort of state which is apt to be induced through perceptual experience and liable to combine with *desire* to generate action; that *pain* is an experience frequently caused by bodily injury or organic malfunction, liable to cause characteristic behavioural manifestations such as groaning, wincing and screaming; and so on.)

Another commonly voiced objection to functionalism was that it is incapable of capturing the felt nature of conscious experience (Block and Fodor, 1972; Nagel 1974). Objectors have urged that one could know everything about the functional role of a mental state and yet still have no inkling as to *what it is like to be in that state* – its so-called *quale* or subjective *feel*. Moreover, some mental states seem to be conceptualized purely in terms of feel; at any rate, with beliefs about causal role taking a secondary position. For example, it seems to be just the feel of pain which is essential to it (Kripke, 1972). We seem to be able to imagine pains which occupy some other causal role; and we can imagine states having the causal role of pain which are not pains (which lack the appropriate kind of feel).

3.3 Theory-theory

In response to such difficulties, many have urged that a better variant of functionalism is *theory-theory* (Lewis, 1970, 1980; Churchland, 1981; Stich, 1983; Fodor 1987). According to this view, mental state concepts (like theoretical concepts in science) get their life and sense from their position in a substantive *theory* of the causal structure and functioning of the mind. To know what a belief is (to grasp the concept of belief) is to know sufficiently much of the theory of mind within which that concept is embedded. All the benefits of analytic functionalism are preserved. But there need be no commitment to the viability of an analytic–synthetic distinction.

What of the point that some mental states can be conceptualized purely or primarily in terms of feel? A theory-theorist can allow that we have *recognitional capacities* for some of the theoretical entities characterized by the theory. (Compare the diagnostician who can recognize a cancer – immediately and without inference – in the blur of an X-ray photograph.) But it can be claimed that the concepts employed in such capacities are also partly characterized by their place in the theory – it is a *recognitional* application of a *theoretical* concept. Moreover, once someone possesses a recognitional concept, there can be nothing to stop them prizing it apart from its surrounding beliefs and theories, to form a concept which is *barely* recognitional. Our hypothesis can be that this is what takes place when people say that it is conceptually possible that there should be pains with quite different causal roles.

Some or other version of theory-theory is now the dominant position in the philosophy of mind (which is not to say that there are no difficulties, and no dissenting voices, of course). And in many of its forms, theory-theory is of-a-piece with the sort of naturalism in philosophy which holds that philosophical and scientific enquiries are continuous with one another. From this perspective, both philosophy of mind and cognitive psychology are engaged in fundamentally the same enterprise – to characterize the nature and operations of the human mind.

4 Developments in developmental psychology

In this section we once again provide a very brisk recent history, this time in respect of developments in developmental psychology.

4.1 The Piagetian account

Piaget claimed that children's initial knowledge of relations of cause and effect is limited to what they see (in this respect his position was close to that of some logical positivists). In his early work he characterized their ideas about causality as restricted by 'syncretism' – in a word, by the tendency to connect everything with everything else (Piaget, 1928, p. 4). If asked to complete the beginning of a sentence such as, 'The man fell off his bicycle because . . . ', children under five or six years will respond with, 'Because he broke his arm' rather than, say, 'Because he lost his balance.' On this basis, Piaget denied that young children are able to detect causal relations. In later work with Inhelder (Inhelder and Piaget, 1958), Piaget reiterated his view that children's scientific cognition is muddled and chaotic, and that their beliefs about events are juxtaposed together instead of causally linked.

Not only did Piaget judge young children to be incapable of identifying causal relations clearly, but he contended that they assign internal states and motives to inanimate objects. For example, they believe that inanimate objects – especially those that move – can possess consciousness and have sensations and emotions just as persons do. Piaget (1928, 1929) interpreted children's answers to questions about the movement and feelings of objects to indicate that their notions of causality are primitive, and reflect an inability to reason about the physical world.

According to the Piagetian analysis, moreover, the causal understanding of young children points to a suite of domain-general processes which underpin the nature of early cognitive development. Children's understanding across domains as diverse as geometry, physics and biology constitute a 'structured whole' in the sense that they share common properties in reasoning and problem-solving. What children know is tied to appearances rather than

involving underlying transformations and causal mechanisms which aren't visible to the eye. In conservation experiments, for example, young children typically believe that when water is poured from a short, wide glass into a tall, narrow one, the amount has changed even though nothing has been added or subtracted. Such children are also supposed not to understand the difference between animate and inanimate objects – believing, for example, that a shadow is a substance emanating from an object but participating with the night; and they attribute animistic qualities to leaves in projecting shadows. Only with increasing age do children know how shadows are projected and deny that objects cast shadows at night. Similarly, Piaget proposed that young children misunderstand the nature of dreams; they believe that dreams originate from outside and remain external to the self.

In later childhood, after the age of seven years, children can use transformations and invisible mechanisms in their causal reasoning. However, not until they achieve a formal operational understanding in early adolescence do they systematically test hypotheses, on a Piagetian approach. While much of this theoretical apparatus has been rejected in later work, many developmental psychologists continue to share Piaget's view that children's understanding undergoes radical conceptual change over the course of development (Carey, 1985; Wellman, 1990; Perner, 1991; Gopnik and Meltzoff, 1997).

4.2 Modern evidence for early causal understanding

We now know that Piaget significantly underestimated children's capacity for causal understanding. Even on the most sympathetic evidence, children's knowledge should be seen as variable rather than as constrained by domain-general stages (Siegler, 1994). It is now well documented that, although no single factor can fully explain children's inability to conserve, their responses on conservation tasks have much to do with the child's perception of the relevance and purpose of the context in which questions are asked (Donaldson, 1978; Siegal, 1997, 1999). Similarly, young children seem surprised when inanimate objects appear to move by themselves, or when they unaccountably appear or disappear (see the reviews by Carey, 2000b, and Rakison and Poulin-Dubois, 2001). Moreover, children as young as three years old have been shown to be very adept in distinguishing real from pretend objects and events in a simplified testing procedure where they are asked to sort items into those which are real (can be seen, touched and acted upon) and those which are not (Wellman and Estes, 1986; Leslie, 1994a).

Young children can also use causal knowledge in the fundamental process of naming and classifying artefacts. In a recent demonstration, Gelman and Bloom (2000) asked preschool children to name a series of simple objects. In