Lucidity and science

II: From acausality illusions and free will to final theories, mathematics, and music

PROFESSOR MICHAEL E. McINTYRE
Centre for Atmospheric Science at the Department of Applied Mathematics and Theoretical Physics, University of Cambridge, UK

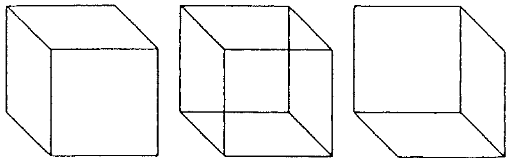
Part I of this series used the idea, more properly hypothesis, that perception works by model fitting. The hypothesis is an important key to clarifying perennial issues about science and the arts and about, for instance, consciousness and free will. What is involved goes beyond what practical people call 'mere semantics' and 'mere philosophy'. For instance, it has practical implications for scientists' professional codes of conduct and for the social experiment we call free market democracy, a theme to be developed in the third and final part of this series. Here in Part II, the model fitting hypothesis is discussed in more detail along with some key evidence. That evidence – much of it checkable by any observant person, with no need for specialist equipment – includes a class of perceptual phenomena to be referred to here as 'acausality illusions', in which, in some cases, perceived times precede the arrival of relevant sensory data. Such phenomena are consistent with the model fitting hypothesis, which predicts that perceived times of outside-world events must be earlier than, and perceived times of internal decisions later than, associated physical events in the nervous system. Associated timespans are typically a few tenths of a second.

A good metaphor for our time, especially for those demanding absolute certainty from scientists, is 'science as our eyes and ears on an uncertain future'. In Part I[72] I gave reasons for thinking that such metaphors are far from superficial. Not only can science be viewed as an extension of ordinary perception – with, of course, suitable precautions and cross-checks – but such an extension is fundamentally what science is. This is because ordinary perception, as well as science, works by fitting models to data from the outside world, of which more in a moment. And if you think that this idea makes science seem less wonderful than you would like it to be, then you may have forgotten how wonderful, how near miraculous, so called ordinary perception is.

Like ordinary perception, science can make certain aspects of the outside world vividly clear to us, and with impressive, indeed awesome, accuracy in some cases. The electron is a sufficient example.[73] Conversely there are, and always will be, aspects that are unclear, illusory, or out of sight altogether, beyond the research frontier. Trying to make sense of things near the research frontier will always be – fundamentally – like driving a vehicle in swirling fog on an unfamiliar, unmapped, twisty road with many branches, and with plenty of oncoming traffic. One has to live with uncertainty, one has to keep one's eyes and ears open, and one has to expect surprises.

Does it make sense to harass, criticise, or sack the driver who admits to uncertainty about what lies ahead? It might be a better idea, in the real world, 'Star Wars' notwithstanding, to sack the driver who shuts his eyes and blocks his ears and claims infallible prior knowledge. This is accurately relevant to public issues involving science. It is relevant for instance to the ongoing fight against new diseases,[74] and – to take an example where I have some specialist knowledge of my own – relevant to the problem of monitoring and predicting environmental change, for instance whether the sea level will rise, and if so how quickly, and by how much, over the next century or two, say by 3 cm or 300 cm, or whether deep ocean circulations could flip chaotically and devastate local climates and economies. For well understood reasons,[75,76] there is great uncertainty over questions like these last two: we are indeed, collectively, driving in the fog, and the 'vehicle' at risk – strong in some ways and fragile in others – is our planetary life support system.

I shall return to such public issues in the third and final part of this series. Here I want to look more closely at the central, all-important, and sometimes controversial idea that both science and ordinary perception work by fitting models to data from the outside world. That science works by such model fitting must be obvious to anyone who knows anything about science. For the moment I shall take this for granted. That ordinary perception works by model fitting is perhaps less obvious, and has been controversial for centuries, perhaps because of the way it seems to contradict subjective experience. For good biological reasons, as recalled in Part I, this model fitting is not only prodigiously fast but also automatic and unconscious. A keen observer will sometimes notice hints of it when the incoming data are sparse, as when straining to see things through fog, or when trying to identify someone's face at a distance. A common experience is the 'perceptual



1 The famous image called the Necker cube (centre figure), to which the visual brain usually fits one of two three-dimensional internal models, one of a cube seen from above, as in the left hand figure, and the other from below as in the right hand figure. If you look at the centre figure for long enough, one internal model tends to be replaced by the other even if you try to stop it happening by an effort of will. The two internal models have comparable prior probabilities and goodness of fit. With an animated, rotating version (available on the Internet[79]), the flipping becomes dramatic: the rotation suddenly reverses

xenomorphosis' in which a familiar face sighted at a distance turns into a stranger's on approach. Before the change, the familiar face is experienced as the perceived reality: the wrong model is fitted at first. This is related to what psychologists call 'categorical perception'. Other examples include the spontaneous flipping of ambiguous drawings like the famous Necker cube (Fig. 1).[77,78] Usually, such a drawing appears to represent one of two 'perceived realities': first you see the one, and then, if you look long enough, the other. Also noteworthy is the 'walking lights' demonstration, described in Part I and available on the Internet[79] – the demonstration that a mere 12 or 13 moving dots, conveying a relatively tiny amount of information, can be enough to generate *unambiguously* the percept of a person walking.

There are far reaching implications, not just for law courts and identity parades, but for understanding most aspects of our existence. Practically everything we do depends on the way perception works; and the way perception works is intimately bound up with the way cognition works. There is no clearcut boundary between the two; and it is no accident that cognition is often called perception, as in stock phrases like 'scientists are perceived as arrogant'. The effect of the walking lights demonstration is not so much to 'see' a person walking but to 'know that you are looking at' a person walking.

The brain must, of course, be using entities that can reasonably be called models, in the sense of being partial and approximate representations of reality. That much is obvious. But the important additional idea is that perception and cognition not only involve using, but also actively building and fitting, internal models or, if you prefer, unconscious symbolic structures:[77] actively fitting hierarchies of model components and subassemblies together in a prodigiously versatile and efficient way, such that relevant parts of the structures are consistent with, and remain consistent with, the relevant sensory data, including data that might be of more than one type, such as visual and auditory – as with the lipreading or 'McGurk' effect, a visually induced change in auditory speech perception,[80] and other multisensory perceptual phenomena.[81–84] The biological point of all this is to have, as far as possible, a single, interconnected, self consistent symbolic structure that when fitted to all available sensory data can represent the surrounding reality, and our place in it, well enough to give us a chance of surviving.

The evidence suggests that the processes involved are not only highly active but also characterisable as self organising, multi-level processes, simultaneously 'top down and bottom up',[80,85,86] that is, with feedback between levels. Such processes are unconsciously 'theory laden', i.e. model dependent; but they are also closely and sharply constrained by sensory data. In a fundamental sense, they are unconscious scientific investigations performed at lighting speed.

The metaphors I am using can hardly be perfect. If there are 'top down and bottom up' aspects then surely there are 'sideways' aspects too, unconscious lateral thinking if you will. The whole cognitive–perceptual complex must indeed be densely and multiply interconnected, or interconnectable, in more than one sense and in many 'dimensions'.[85] However, the basic idea, that both science and ordinary perception work by model fitting, is not only a simple idea in itself but also helpful, I shall argue, toward clarifying many otherwise perplexing problems – far more than how to write and speak lucidly, my starting point on this journey.

As already hinted in Part I, the model fitting idea is helpful toward understanding, for instance, the very origin of our symbolic and linguistic cognitive abilities – unconscious as well as conscious – including, I shall argue, the cognitive abilities associated with mathematics and music, more fundamental to our nature than is often supposed. The model fitting idea contains important clues as to why some educational methods work better than others. It is basic to understanding the phenomena called free will and consciousness. And, as I shall try to show in Part III, it can help to blow away the philosophical confusions about science now threatening the stability of democratic nations; in particular, it offers hope of reconciliation in the current and deeply damaging 'science wars'.

All this deserves serious discussion, more serious and more careful than the foregoing. What follows is an attempt toward providing such a discussion, on the basis of key evidence and in a way that avoids getting too technical.

The model fitting hypothesis

Though hardly novel, the idea, more properly hypothesis, that perception works by model fitting is bold – some would say brazen – not only because of the necessary complexity and interconnectedness of the internal models but also because we have so little detailed physical, chemical, biological, and computational knowledge of how the model fitting might work. We have only the merest glimpses of how neurons, synapses, protein molecules, and other elements in the brain might carry out perceptual processing, and of how its mechanisms develop in infancy and childhood through the intricate interplay of so called nature and nurture. All that can safely be assumed about such details is that the usual textbook models of brain function do not adequately represent them. Such textbook models, which idealise neurons or nerve cells as simple logic elements, undoubtedly describe important aspects of brain function. But there are a billion or more protein molecules, all in thermal motion on picosecond time-scales, within each single one of our tens of billions of neurons.[86] Many of these protein molecules are themselves precisely functional logic elements. So it would be more accurate to view the brain as a hypermassively parallel computational system of which neurons and their synaptic interconnections are, so to speak, merely the tip of the iceberg.[87]

But, as already suggested, the case can be made independently of detailed mechanisms. The basis on which support for the model fitting hypothesis is claimed here is the usual scientific basis, already touched on in Part I. I shall return to this crucial point ('On epistemology' etc., below), but, in brief, if we hypothesise that perception works by model fitting then we can easily explain, and greatly clarify, in a coherent and self consistent way, a host of perceptual phenomena that would otherwise be incomprehensible. Just how completely incomprehensible is well brought out by, for instance, the historical and philosophical sketch and the scientific discussions in Chap. 14 of Ref. 78. The same point is underlined by the celebrated intellectual struggles of the great philosophers themselves.

The model fitting hypothesis is also in keeping with general knowledge about biological systems, and with detailed knowledge at molecular level in some cases. One such case is that of a mammalian immune system such as our own, which can be thought of as a perceiving system – a system that perceives and knows about invading foreign material, for instance detailed molecular patterns on the surface of a virus or bacterium. 'Perceive' is a better term than 'recognise' because the immune system can do it with wholly novel patterns, chemically synthesised patterns that were non-existent until human chemists made them. The system works by actively constructing, from pre-existing components, its own approximate internal models of the incoming patterns.[88] Models are selected from a combinatorially large ensemble on the basis of goodness of fit. The selection mechanism is a clearcut Darwinian mechanism. The information defining the models thus selected is then stored within special memory cells. Here, despite massive complexity, the mechanisms are known in considerable detail at molecular level; and they include a means of iteratively tightening the fit by successive approximations.[89] The whole process has the general character I am talking about, and whose wider significance I am trying to suggest: it is hypermassively parallel, self organising, and simultaneously 'top down and bottom up'. This general character seems typical of biological systems;[87] and there seems no reason why the brain should be an exception, its vastly greater complexity notwithstanding.

Rather than trying to discuss all the relevant classes of perceptual phenomena in what follows, I shall single out just two, the first involving hallucinatory phenomena and the second involving what I shall call 'acausality illusions'. Acausality illusions seem less widely known than their importance warrants, and will be discussed in some detail.

I am not claiming, by the way, that any of the points made here are original. Most if not all of them can be found in a widely scattered literature, though not always interrelated nor put in the simplest possible terms. Ideas equivalent to, or foreshadowing, the model fitting hypothesis for perception can be found in very many publications, going back for instance to Berkeley and Kant[78] and perhaps even, stretching a point, to some of the Ancient Greeks. The Internet version of this article[79] lists over 30 references within a far vaster literature, including those mentioned in Notes 5–7, 22–38, and 45–47 of Part I. Sometimes the idea that there are mental models is made use of without recognising its full generality, nor its full significance. References 45–47 of Part I trace some of the more recent philosophical history back by a century or more, to the time of Charles Darwin, and show in more detail why the model fitting hypothesis has been controversial and to some extent still is. My purpose is merely to bring out the simplicity, accessibility, and checkability of the main points and to argue, from the experimental evidence, that there is little room for controversy today. The evidence is plentiful, important parts of it can easily be verified by any observant person, and there are very many consistency checks. The discussion will lead us naturally toward the wider implications already mentioned, practical as well as philosophical, including the implications for the public understanding of science.

Hallucinations, 1–1 mapping fallacies, writing, and education

The existence of perceptual experience in the absence of sensory data, or in conflict with sensory data, is one of the plainest and most direct lines of evidence in support of the model fitting hypothesis. There are ordinary nocturnal dreams and there are waking hallucinations, occasionally having the full force of perceived reality.[90–92] To anyone who discounts the traditional, magical or demonological, 'explanations', such phenomena are incomprehensible – practically impossible to explain – without hypothesising the existence of internal models of some kind.

One such case is described in detail by the neurologist Oliver Sacks.[92] It is the case of the painter Franco Magnani, whose work was exhibited at the

San Francisco Exploratorium in 1988. The paintings all depicted Magnani's childhood village of Pontito, Tuscany, which at the time of the exhibition he 'had not seen for more than thirty years'. Some were painted as if from 'an imaginary aerial viewpoint fifty or five hundred feet above the ground', to which he could not have had access in reality, ruling out simplistic photographic memory hypotheses. The exhibition compared the paintings to photographs taken after the paintings were done, including photographs from 'a camera aloft on a pole'. 'It was as if Magnani held in his head an infinitely detailed three dimensional model of his village, which he could turn around and examine, or explore mentally, and then reproduce on canvas with total fidelity.' Magnani himself, in other respects an ordinary human being, reported waking 'visions' in which he saw, heard, smelt, and *felt* countless aspects of the village 'with a minute and three dimensional quality that he compares to holography'.

Such evidence is gradually helping to dislodge a philosophical tradition, still influential today, that begs all the questions by presuming ordinary perception to be what it seems to be subjectively: a cameralike or microphonelike process, a passive 1–1 (one to one) mapping or veridical imprinting of so called sense 'data', a sharp and straightforward correspondence between appearance and reality. This traditional presumption could be called the *1–1 mapping fallacy*, or *veridical perception fallacy*, or *cat on the mat fallacy*. It underlies some of the difficulties in the philosophy of science;[93] and Sacks[92] notes how persistently its influence has impeded the recognition and understanding of non-trivial perceptual phenomena and their clinical pathologies, sometimes with tragic personal consequences.[94]

There is a similar fallacy that probably contributes to the expertise in bad writing referred to in Part I. It could be called the *1–1 lexical fallacy*, or *explication-of-concepts fallacy*. This is the tacit presumption of a fixed, sharply defined 1–1 mapping between individual words or phrases (lexemes) on the one hand, allowing for synonym groups, and the things they refer to on the other, implying a correspondingly unimportant role for context – as if context and word patterns were somehow secondary. Exactly such a fallacy led to the early fiascos in language translation by computer, and probably underlies, also, the feeling that writers can safely indulge in the gratuitous variation of words and word patterns, as if intelligibility were merely a matter of recognising synonyms, a simple matter of dictionary lookup and nothing else. How profoundly wrong this is can be seen from the experience of learning to write, and from the evidence and the careful discussions in, for instance, Ref. 85 and in Refs. 6, 12, 45, 62, and 68 of Part I. As a simple illustration, consider French sentences of the type 'Je ne sais rien' and their counterparts in some dialects of English, 'I don't know nothin'. These illustrate how human language works with perceptually grouped, syntactically functional word patterns, ahead of conscious thought, rather than working with isolated words or phrases. The one thing our language instinct does not do, whatever the 'Procrustean grammarians' may say, is to apply rules of logic at the superficial level of word-by-word dictionary lookup.

One may surmise that 1–1 mapping fallacies also underlie a general misconception that often seems to confuse the politics of education. There seems to be a strongly held lay view of education itself as a process of 1–1 mapping or passive imprinting, to be most efficiently implemented by intimidation and rote learning. This is cognate with the cultural relativist *blank slate* or *tabula rasa*[78] fallacy.

Acausality illusions: the theoretical possibility

A further line of evidence is the existence of acausality illusions, in which effects seem to precede causes, usually by some fraction of a second. They were perhaps first noticed by speech perception researchers,[95] and awareness of them has recently begun to clarify discussions of free will and consciousness, as noted below. They have sometimes been called 'after going effects', or 'backward referral in time', and put forward as posing a deep mystery – perhaps demanding a solution to the entire 'problem of consciousness'[96,97] and perhaps, it has even been suggested, overthrowing elementary and thoroughly checked principles of physics as well. Here I argue that, on the contrary, acausality illusions are indeed illusions, and that to understand them we do not need a full understanding of brain function and consciousness, let alone a revolution in physics. On the contrary, I argue, acausality illusions are robust perceptual phenomena, easy to study experimentally, and straightforward to explain by the model fitting hypothesis. I would also argue that their existence is incomprehensible without some such hypothesis.

The existence of acausality illusions provides, in particular, a very direct refutation of the 1–1 mapping fallacy. When applied to the time domain, 1–1 mapping says that perceived times must always be in 1–1 correspondence with the temporal order of events in the sensory data. But that is contrary to what is found experimentally, especially over timespans less than half a second or so; and it is something that anyone can check for themselves.

Before turning to examples let us note why something of the sort has to be expected. The reason is clear from a biological viewpoint.[98–100] Perceived times must make allowance for finite cortical processing timespans in fast moving survival situations like hunting and fighting. As with today's sports, such situations demand eye–body coordination over timespans of the order of milliseconds. In some ball games the ball can travel several centimetres in one millisecond. The best ball players must therefore, like the best musicians, be timing their actions to within a small number of milliseconds. Such timespans are around a hundredth of the associated cortical processing timespans. Many stages and levels of cortical processing must be presumed to be involved in generating a conscious percept from, say, visual and auditory data, such as the perceived time of hitting a ball or of sounding a note on the piano; and the psychophysical and electrophysiological evidence indicates typical processing timespans of the order of hundreds of milliseconds.[101] The implication, counterintuitive yet inescapable, is that perceived times, in order to be useful when coordinating an action like hitting a ball, must not only be accurate to milliseconds but must also be able to *precede* some of the cortical processing involved. If something is perceived to happen at 12 noon minus, say, 100 or 200 ms, as with a piano note struck just before the 12 noon time pip heard on the radio,[102] then some of the cortical processing required to generate the percept is likely to have taken place after 12 noon.

There is no conflict with ordinary physical causality, i.e. with the well checked principle that physical causes precede physical effects, if perceived times are properties of an internal model. There is then no need for perceived times to coincide with, for instance, the physical times of any cortical processing events involved in the construction, testing, and updating of that model, nor with the physical times of any other events in the nervous system. There is then no mystery about the exquisitely precise performances of the most skilful musicians and sportspeople, despite cortical processing timespans of hundreds of milliseconds during which a ball might travel tens of metres. Rather, that precision means just two things. First, there is an internal model that has goodness of fit, to data from the outside world, at time resolutions of the order of milliseconds in some cases. Second, the internal model has predictive power over timespans of hundreds of milliseconds or more. It must have at least that much predictive power in order to be useful. These are the same goodness of fit and predictive power that keep us on the road when driving a vehicle at high speed in good visibility. The predictive power is deployed automatically, involuntarily, and ahead of conscious thought, something that becomes noticeable after narrowly avoiding a high speed collision. One observes oneself avoiding the collision, but the action takes place ahead of conscious thought.

Once these simple though counterintuitive points have been grasped – of which the most important is that a perceived time, being a property of an internal model, need not coincide with the physical time of any cortical or other neural event – it becomes obvious that perceived times can, at least as a theoretical possibility, precede even the arrival of relevant sensory data. Much confusion has arisen from tacitly ignoring this possibility. And, as already indicated, the possibility is realised. Perceived times do indeed, in many circumstances, precede the arrival of relevant sensory data. This is the most easily studied type of acausality illusion. Examples of such illusions are accessible to any careful observer. There are examples from speech and music perception, from tactile perception, and from vision.[95,99,100,103,104]

Visual acausality illusions

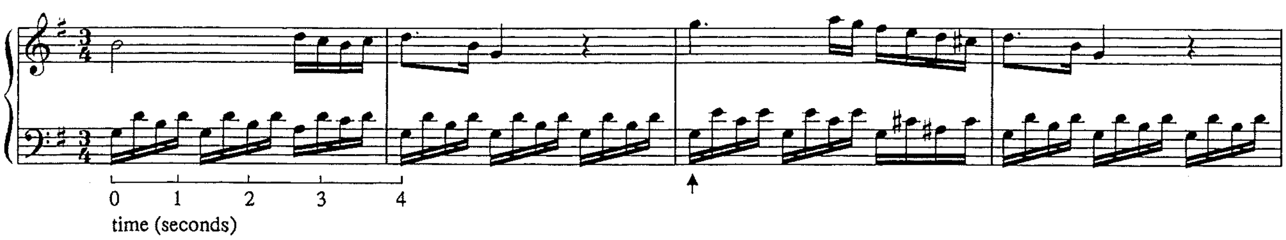
The visual examples include the classic apparent-motion illusions (*see* Refs. 78, 80, and 99 and references therein). The simplest case is what one sees when a small, stationary object in the field of view disappears and then reappears in a neighbouring position. Under suitable conditions and with suitable timing, over timespans, again, of a few hundred milliseconds, one sees an illusory motion. The object appears to move away from the old position toward the new, and to begin to do so before the arrival of the light signal carrying information about the new position. The effect is familiar from certain kinds of illuminated advertising displays, and from the warning lights, a side-by-side pair flashing alternately, used in some countries to warn motorists that a railway train is about to cross the road at a level crossing. As the designers of such warning lights must have realised long ago, the apparent motion attracts attention more powerfully than the flashing of a single, stationary light, and is well worth the extra cost.

Such apparent motion is a robust, repeatable perceptual phenomenon that anyone can investigate for themselves, with commonly available equipment. Nicholas Pinhey and I have investigated it with a computer driven cathode ray screen. The screen showed a bright disc about a centimetre in diameter on a dark background, at viewing distances of the order of half a metre to several metres. The exact colours, dimensions, and viewing distances were not critical. The disc was displayed for 200 ms in one position and then, after a 200 ms pause with the screen blank, displayed for another 200 ms in another position several centimetres to the right. Then, after another 200 ms pause with the screen blank, the whole cycle was repeated a number of times.

In the clearest version of this demonstration, among those we tried, the time of appearance of the disc in its right hand position was emphasised by a bright flash, created by making the disc white for the first 20 ms then red for the remaining 180 ms. In its left hand position the disc was red for the full 200 ms; there was no white flash. With this display, the apparent rightward motion plainly began before the perceived time of the flash.[105,106] The only assumption made here, which could be criticised but which I believe will stand scrutiny, for the reasons already discussed, is that any mismatch between the perceived and physical times of the flash is much less than 200 ms. We know that it must be much less in other cases, such as the perceived time of hitting a ball. A ball travelling 3 cm in 1 ms travels 6 m, nearly 20 feet, in 200 ms.

← Demonstration at www.atm.damtp.cam.ac.uk/people/mem/papers/LHCE/

Our usual subjective experience of perception, as not only passive but also instantaneous – our



2 Opening bars of the slow movement of Mozart's C major piano sonata, K.545

subjective unawareness that perceptual processing, by all the evidence, takes at least hundreds of milliseconds – could itself be taken as an example of an acausality illusion.

Auditory acausality illusions

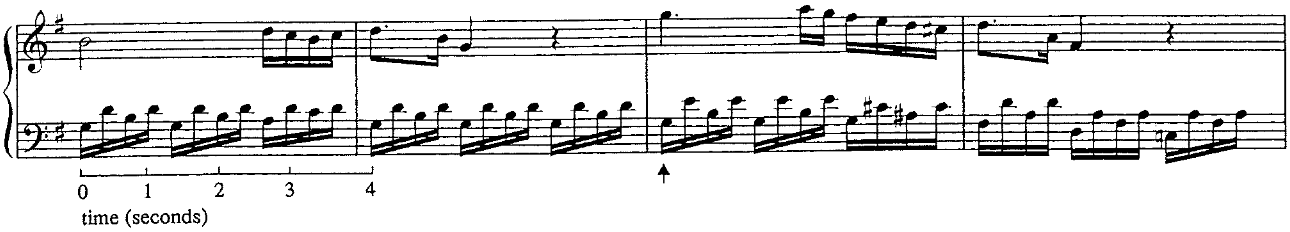
Music, 'the art that is made out of time',[107] gives us outstandingly clear and simple auditory examples. These deserve careful discussion. Consider for instance the change of harmony at the third bar, shown by an arrow, in the Mozart example reproduced in Fig. 2 (audio available on the Internet[108]). Even at the fairly slow tempo shown by the time markers, the harmony change is perceived to occur at the time of the arrow, even though the information defining the change is entirely contained in the following two notes, the first of which does not sound until one-third of a second later. This acausality illusion depends only on the listener's general familiarity with the 'language' of Western tonal music. It does not depend on the listener's having previously heard the particular piece of music, any more than understanding a sentence of speech depends on having previously heard the particular sentence. The perceived time of the harmony change is especially clear to a musically trained listener because, as with sports training, musical training cultivates an acute awareness of perceived times through the need to monitor and control timings in performance.

Once again we are dealing with perceptual grouping.[77,80] Recall again the simplest visual examples, such as that in Fig. 3 below illustrating grouping in fours. In the auditory example of Fig. 2, the notes of the lower part are also grouped in fours, and perceived as coherent units.[80] In particular, the four notes starting at the arrow are grouped together, as might be expected from their significance for the harmony. The same grouping is encouraged by the invariant qualitative shape of each foursome within an organically changing pattern-sequence. Such grouping phenomena are essential, also, to speech perception[103] – compare the spoken sentence 'I'll get the ten to ten train' with 'I'll get the tentative plan checked'. The perceived timing of the harmony change at, not after, the time of the arrow in Fig. 2 is confirmed by the way in which a composer or arranger skilled in Western tonal music would devise an orchestral accompaniment for this example, if intending to preserve its musical sense. Such an accompaniment would change harmony at, not after, the time of the arrow. This is to be contrasted with the way the accompaniment would need to be written if musical perception worked by temporal 1–1 mapping. The accompaniment would then have corresponding delays in the harmony changes. Audio clips and full scores to illustrate both kinds of accompaniment are available on the Internet, for anyone who would like to hear the effects for themselves.[108] With the 1–1 mapping version, the bar starting with the arrow sounds unlike anything that I have ever heard in Mozart's music, discounting accidents of performance. In the subsequent bar, the musical sense is changed very strongly.

3 Visual grouping in fours

If you happen to be familiar with the particular Mozart piano solo quoted in Fig. 2, and think that such familiarity might be influencing how you hear things at the time of the arrow, then try listening instead to the the piano solo shown in Fig. 4 (audio again available on the Internet[108]). There is a new harmony change, strongly audible and very different from Mozart's. But its perceived timing still corresponds to the time of the arrow, even though the example is identical to Mozart's for a full two-thirds of a second beyond the arrow. A suitable orchestral accompaniment (available on the Internet[108]) would again change harmony at, not after, the time of the arrow.

Timespans of two-thirds of a second are by no means the longest that can be claimed for musical acausality illusions.[104] But two-thirds of a second is already interestingly different from one-third of a second: two-thirds of a second seems to be long enough for the updating of the internal model to be made consciously noticeable through a sufficiently strenuous concentration of attention. At least it can be made thus noticeable to myself. Individuals may well vary. By listening to the piano solo of Fig. 4 while concentrating hard on the instant marked by the arrow, I think I can catch an audible, though fleeting, impression of Mozart's version being superseded by a strong perception – a perception having the subjective force of actual experience, of unique 'perceived reality' – that the harmony change at the arrow was, 'in fact', the new one. This is an example of what Dennett,[99] in his interesting discussion of these matters, would call a genuinely 'Orwellian' or revisionist phenomenon. My experience with the last



4 This piano solo is exactly the same as Mozart's until the second note after the arrow

example also seems consistent with Dennett's remarks about subjectively probing perceptual content in different ways, also called 'top down' effects in the literature, in the sense now of being consciously driven. (This again is where one-dimensional metaphors like 'top', 'high', 'deep', 'multi-level', etc., begin to show their limitations.) When listening to the last example as music, in the ordinary way, I do not hear even a fleeting impression of Mozart's version. I hear only the new harmony change at the arrow, even though I may well imagine, not hear, Mozart's version in a mental comparison.

The subjectively noticeable difference between two-thirds and one-third of a second is consistent with available information about cortical processing timespans. In ensemble averaged, event related electroencephalograms,[101] for instance, the recorded activity often has a strong peak near one-third of a second, and weakens thereafter.

Consciousness, and the unconscious drive to prune

All the perceptual phenomena described above are easily accommodated by the model fitting hypothesis, which requires that internal models be continually updated, on penalty of losing goodness of fit and any chance of useful predictive power. As with the avoidance of a high speed collision on the road, the updating of an internal model has to take place whether or not the updating is consciously noticeable as such. As has been pointed out in the literature,[99,103,104,109] the updating must involve pruning a tree of model possibilities that extends forward into the future at any given instant, each forward pointing branch having its own prior probability and predictive implications. The set of model possibilities needs to have some such branching structure, in order to allow the evolving internal model to be efficiently and rapidly updateable by parallel processing. Dennett[99] aptly calls this a 'multiple drafts' structure. Because of finite cortical processing timespans, the branching of model possibilities must extend forward not from the present but from times hundreds of milliseconds in the past or earlier.

We are normally conscious of one branch only, the end result of the pruning, for an obvious reason: to be *conscious* of more than one would be fatally confusing in a survival situation.

The unconscious drive to prune to just one branch, to perceive a unique reality, to fit a single model, is therefore overwhelmingly strong. In the absence of sufficient data the pruning tends to take place anyway, using prior probabilities. It is easy to find musical examples in which the harmonic motion is heard as unambiguous even when the scoring is sparse and not all the notes of the perceived harmony are sounded; Mozart's famous string serenade 'Eine Kleine Nachtmusik', K.525, has a conspicuous example halfway through the second full bar of the slow movement, at the high point of the melody. The corresponding point in speech perception – no perceived ambiguity despite substantial actual ambiguity – is well illustrated by the work of Swinney, Seidenberg, Tanenhaus, and others (Ref. 104, p. 144), and in vision it has already been illustrated by the walking lights demonstration,[79] and by the Matisse drawing in Part I. *See* also, for instance, Bregman's[80] insightful remarks about the 'all or nothing' character of the internal model fitting process, as indicated by psychoacoustic experimental data. The phenomena called 'cognitive illusion', 'mental tunnel', and 'mind set', even if involving other, 'more cognitive' aspects of brain function, illustrate what seems to be fundamentally the same point.

What is remarkable is not so much the tree pruning but, on the contrary, having even the slightest ability to postpone it, even in the 'more cognitive' parts of our minds – to be able to stay conscious of multiple branches, of multiple possibilities, for any purpose whatever. The ability thus to live with uncertainty, an important part of our adaptability, could be one of the biological reasons for consciousness itself. Nevertheless, staying conscious of three or more possibilities is more difficult, and probably more energy consuming, than staying conscious of two. There is a noticeable human tendency toward 'false dichotomisation', the urge to prune to just two simplified, mutually exclusive branches or possibilities even when it makes little sense to do so – as with 'capitalism or socialism', 'science or religion', 'nature or nurture', and 'strong Earth or fragile Earth'.[110] There are further biological reasons for this tendency toward false dichotomisation, beyond the mere saving of energy, to be discussed in Part III.

Could the list of acausality illusions include the effects that sometimes seem noticeable in ordinary conversation? 'You interrupted me!' 'No, *you* interrupted *me*!' The model fitting hypothesis says that it is possible, in principle, for each person genuinely to perceive – with the subjective force of actual experience, of unique reality – the other's speech as starting a small fraction of a second later than their own. Using a tape recorded dinner table conversation, I have checked that this has indeed happened to me: I experienced my speech as starting fractionally before another's, but the tape showed the reverse. In such cases, by contrast with cases of eye–body coordination in hunting and fighting, the accuracy of perceived timings would seem less important for survival; so perceived timings in conversations might fit acoustic events less precisely.

I have not, by the way, seen the term 'acausality illusion' used in the psychological or philosophical literature, but I dare to use it here because it seems both simple and apt, emphasising as it does that nothing mysterious is involved – that there is only an apparent, not a real, conflict with physical causality. As already suggested, acausality illusions exist for a clear biological reason, the need to grasp space and time together, to have an accuracy in perceived times commensurate with the accuracy in perceived positions of fast moving objects. In fast moving survival situations, spatial accuracy would have little value without a corresponding temporal accuracy. The spatiotemporal character of internal models must, indeed, be biologically ancient, almost as ancient as predators and prey. Seen in this light, the existence of temporal as well as spatial perceptual illusions is not surprising.

Consciousness and free will

The biological need to grasp space and time together includes the need to coordinate internal decisions with external events. So, again not surprisingly, there is a second kind of acausality illusion, which concerns the perceived time of taking a decision to act. This point has been overlooked in some of the debates about consciousness and free will.[97] Perhaps the most striking example, with the clearest experimental evidence, is the acausality illusion evoked in the slide projector experiment of Grey Walter. This was first described in 1963 in an unpublished report, summarised in Refs. 99 and 100 and more recently discussed, along with related phenomena and philosophical issues, in Refs. 111. Neurosurgical patients were invited to entertain themselves to a slide show by pressing a button to advance the slide projector, at times of their choosing. But the projector was wired not to the button but directly to a certain part of the patient's motor cortex; and the subjective effect – startling and disconcerting to the patients, who must have wondered whether they were going crazy – was that the projector seemed to behave acausally, to anticipate their decisions. The projector seemed to advance itself just before they decided or, rather, perceived themselves as deciding, to press the button.

This too is easily accommodated by the model fitting hypothesis, which says that perceived times, including perceived times of taking decisions, and perceived times of pressing buttons, are properties of the brain's internal models rather than the physical times of any associated cortical activity. It also underlines the point about free will. The existence of acausality illusions implies that questions about the perceived times of decisions are separate from questions about whether the decisions were taken freely or not. There was no reason, in this experiment, to doubt the patients' freedom to decide to press the button, where freedom is understood in the ordinary, everyday sense related to personal responsibility. The experiments add to our knowledge about perceived times and acausality illusions, but say nothing at all about personal freedom of choice, and personal reponsibility for our actions. They do not say, as has sometimes been argued, that free will is illusory. What is illusory is the perceived time of willing the action.

In principle, the experiment could be repeated with a gun instead of a slide projector. The point is that questions about whether the patient intended to murder someone are separate from – are nothing to do with – questions about the precise timing of the murder, about whether the bullet struck the victim a fraction of a second earlier than intended.

The model fitting hypothesis says that the conscious self, the perceived self that I experience as having intentionality, as being free to make choices, as planning things and taking decisions, must, like any other percept, arise from an internal model: the single internal model that my brain fits to my real, biologically diverse, multi-component, multifariously subtle self.[98,99,112] As Ref. 112 aptly puts it, the brain viewed from outside looks more like a committee, composed of different parts that evolved at different times for different purposes (of which we get occasional hints: 'something tells me that ...', 'my head says one thing but my heart another', and so on). Yet[112] 'I am utterly convinced that there is only one me ... not some kind of committee.' The model fitting hypothesis makes sense of this paradox: the brain has multifarious parts, but only one self-model.

However diverse our internal makeup may be, in order to survive we need, continually, to make sense of our surroundings and our own location and orientation in those surroundings. So the repertoire of internal models and submodels that are used to construct the perceived world, with its stationary and moving objects, has to include a self-model. Simultaneously with other models, this has to be fitted to the incoming sensory data including, now, internal data from one's own body, such as proprioceptive data about limb positions. The end result is a single spatiotemporal model of oneself in one's surroundings. If this model fitting process fails, one may become 'disoriented'. Because of the need to coordinate internal decisions with external events in fast moving situations, the model property called perceived time must be a single property, defined consistently, of the entire model of oneself in one's surroundings. It is this single model property that represents not only the 'when' of when a ball is hit, or a piano key struck, but also the 'when' of taking decisions or initiating action. Because cortical processing is necessary to arrive at even a snap decision, the perceived 'when' of such a decision must, inevitably, be preceded by cortical activity, such as that causing Grey Walter's slide projector to advance.

In summary, then, the model fitting hypothesis predicts that perceived times of internal decisions must be later than, and perceived times of outside-world events earlier than, at least some of the associated physical events in the nervous system. Only thus can the brain, with its finite rate of information processing, typically taking hundreds of milliseconds, consistently represent both sets of times in its internal model of the self in its surroundings at the far finer, millisecond, accuracies needed for survival. Free will and intentionality are properties of the self-model.

On epistemology and final theories: model fitting is the best we can do

The reader will have noticed that the entire discussion so far, including the discussion of perceptual phenomena, has been based on the usual working scientist's adherence to what I called the scientific ideal, giving primacy to coherence (including self consistency) and goodness of fit of models to data. In particular, the hypothesis that perception works by unconscious model fitting is itself a model, as yet incomplete and sketchy: a consciously constructed model, a scientific theory, though still a rudimentary one, to be tested against relevant data. It has passed many such tests including those discussed here (and *see*, for example, Refs. 80–84 and 92, among many others), and it has no serious rivals.

Of course the testing of the model fitting hypothesis involves treating as data what other people say or write about what they do, see, hear, smell, and feel, especially experimental psychologists and their subjects. I am being consistent in accepting this precisely because, like most scientists and some philosophers, I respect the scientific ideal in the form adopted here. This means that I am prepared to do the best I can with repeated trial and error plus multiple consistency checks, rather than looking for anything like absolute proof, seeing the latter as unattainable;[113,114] and I expect to find errors in the data as well as in the model building and model fitting, relying on the consistency checks to detect most errors sooner or later. This attitude is no different, in principle, from that adopted in other scientific investigations. Precautions against human wishful thinking and deceit are no different, in principle, from precautions against laboratory contamination.

It follows that I am rejecting, for instance, certain behaviourist and positivist views of so called scientific method. These forbid treating as scientific data what people say about their subjective experiences. We can reject such views if only because they are self inconsistent, hence self refuting: their adherents expect scientific colleagues to take seriously what they themselves say they did, saw, heard, and felt. Underlying this superficial inconsistency is a deeper inconsistency, which replaces cautious respect for good experiments and data by a tacit assumption that observational or experimental 'facts' are perceptible in some absolute sense – unambiguously and directly, with no dependence on conscious or unconscious model fitting hence no dependence on unconscious modelling assumptions – and that only such 'directly perceived facts' are to be taken note of, and regarded as 'real', by the rigorously objective scientist. This means first that the veridical perception fallacy is built in from the start, and second that telling anyone else about what was perceived is strictly speaking futile, making the intended objectivity hardly distinguishable, in principle, from the ultimate subjectivity of solipsism.

Positivism even goes to the extreme of claiming that 'observed facts' are the only meaningful entities. Positivism was historically important as a stepping stone to relativity and quantum mechanics, at a time when trying to think about 'reality' was becoming impracticably difficult; but its oversimplifications and inconsistencies are now increasingly recognised as such by serious thinkers.[115,116] One might characterise the positivist viewpoint as putting an apt emphasis on science as model fitting, while, with breathtaking inconsistency, saying that, apart from the aspects we 'perceive directly', we can never claim that the models represent reality. To put it more sharply, and to underline the inconsistency, positivism says that consciously fitted models cannot claim to represent reality but unconsciously fitted models can.

Such absurdities and dead ends are avoided by the scientific ideal in the form adopted here. The ideal avoids other philosophical complications as well, such as infinite regressions of $(\text{meta})^n$-languages and $(\text{meta})^n$-models. It does all this by acknowledging what working scientists are familiar with from experience, that science works by model fitting and that we therefore cannot do better than repeatedly checking for coherence, self consistency, and goodness of fit in as many ways as possible, and using Occam's razor, or explanatory parsimony, to mitigate the combinatorial explosion of model possibilities. There is a principle of humility or, if you will, an 'epistemological uncertainty principle' that is useful to recognise. This accepts that there is no way of proving anything absolutely, even the existence and uniqueness of the outside world. It accepts that the best we can do is to adopt coherence, self consistency, and goodness of fit, supplemented by Occam's razor, as the most central and fundamental requirements.

Humility does not necessarily mean being apologetic. The best we can do may be very good indeed. The requirements are stringent. To be sure, they permit non-uniqueness, what Kuhn and Feyerabend[93] have rather sweepingly called the 'incommensurability' of scientific theories. More than one simple model might fit the data adequately. But the requirements do not permit the kind of arbitrariness that has been read into this by, for instance, some of the cultural relativist and social constructivist philosophies and by the journalists they have influenced, the idea that 'theory laden observation' means seeing what you like. Scientific experience argues strongly against such arbitrariness, just as does ordinary experience, as when driving in fast traffic. No-one would claim that, under conditions of good visibility, the perceived positions and velocities of vehicles are arbitrary, any more than that they are absolutely and perfectly accurate.

The requirements imposed by the scientific ideal mean that, for a scientist, or for anyone who respects the ideal, to change one's mind when necessary is a virtue and not a fault. Making this clear seems to be a perennial difficulty in exercises to improve the public understanding of science; and the difficulty is worsened by the mythical, quasi-religious belief that science is about discovering absolute, final, and infallible truth, of which more in Part III. Science is, or should be, self-correcting:[90] change is necessary whenever inconsistencies emerge, as may happen when new experimental data become available. When a driver on a foggy road catches glimpses of what might or might not be an oncoming vehicle, it is a virtue and not a fault for the internal model in the driver's brain to change, at least provisionally. In the scientific counterpart of this, 'internal model' has to be understood in a broad sense that includes not only scientific 'theory' but also experimental concepts and technicalities. As emphasised in Part I, the scientific ideal demands that the whole conceptual edifice, 'experimental' and 'theoretical', should be self consistent – never a trivial matter.

For example, the significance of the proverbial 'pointer readings' of experimental physics, more likely nowadays to be groups of pixels on a computer display, is not that they represent absolute, directly perceived 'facts' independent of, for example, the experimenter's conceptual models of how the apparatus works. Rather, the pointer readings or pixel intensities provide some of the sensory data that have to fit into the whole conceptual edifice in ways that withstand consistency checking by, for instance, a second experimenter looking at the same apparatus and thinking about how it works, and about what the first experimenter thinks he or she is measuring.

Philosophically tricky though it may be, I can also reasonably claim – without having to enter the mind–brain debate – to be consistent, for present purposes, in having stretched the concept of 'outside world' to include aspects of my own brain function. A sufficient justification is that I am avoiding the behaviourist and solipsist views and can therefore admit that my brain function could be investigated by any scientist. Being one of Grey Walter's neurosurgical patients need not stop me from accepting his experimental results. At least it need not do so unless an inconsistency were to be found, such as my not feeling startled when he said I behaved as if startled. The scientific ideal, in the form I have stated it, copes well with this problem even when the term 'outside world' becomes inadequate. (One can switch to other accepted usages such as 'real world', 'objective world', or 'physical world', at the risk of bringing in other prejudices.) Anyone with philosophical worries about such issues can find careful and insightful discussions in Refs. 98 and 99, the latter under the heading 'heterophenomenology'. But the philosophical problems are most simply and quickly circumvented, it still seems to me, by the scientific ideal in the form adopted here, which emphasises consistency checking and goodness of fit and avoids, from the outset, prior judgments and taboos about detailed methodologies and domains of applicability.

Finally, mathematicians, logicians, and other careful thinkers may well feel uncomfortable about giving a more fundamental status to consistency checking, which can include circular chains of reasoning, than to time honoured, economical, powerful, and elegant serial reasoning. I share this discomfort; but it is clearly unavoidable. No knowledge about the 'outside' world, the real world, the physical world, physico-chemico-biological reality, whatever you want to call it – nor even any assurance of the existence of that world – can possibly be deduced from serial reasoning alone. Serial reasoning has to start somewhere, but there is no guaranteed starting place. We are not given any self consistent set of axioms from which we are guaranteed beforehand to be able to deduce anything, let alone everything, about the world. To be sure, the existence, in principle, of such a set of axioms is tacitly assumed, or could reasonably be said to be assumed, by advocates of the famous but putative Theory of Everything, or Final Theory of the world or universe. More precisely, the assumption implicit in the words 'everything' and 'final' as usually understood – and as would surely be understood by most lay people – is that one can have a self consistent set of axioms, prescribed in advance, from which everything that can be said about the world, such as the probability of radioactive decay of an atomic nucleus, can, in principle, be deduced by explicit, rigorous rules, also prescribed in advance. But that is exactly what is known, from Gödel's theorem,[96] not only to be impracticable but also *impossible* in principle if the set of axioms is rich enough for the purpose.

Indeed, and more clearly to the point, Gödel's theorem tells us that we cannot even know, from the axioms and rules of deduction alone, whether or not those axioms and rules are self consistent. This in itself – leaving aside the question of how goodness of fit is to be finally, absolutely, and infallibly tested – should be enough to warn us that belief in the existence of a Final Theory, even in the weak sense of being knowably the best fundamental theory that can ever be found by humans, is strictly a matter of faith. Stronger beliefs require stronger faiths; and I return to this point in Part III because of its political and humanitarian importance.

Steven Weinberg, perhaps the most impressive, aesthetically compelling, and intellectually powerful advocate of something approaching belief in a complete and final theory of the universe – not just a major breakthrough but the final breakthrough – candidly and honestly admits that such belief is indeed no more than a matter of faith,[115] and moreover that today's hopes of finding such a theory rest on a further article of faith, the very shaky assumption that today's quantum mechanical principles are themselves complete and final. This is in fascinating contrast with what, in my opinion, is Weinberg's most deeply perceptive remark, following Bohr (Ref. 115, p. 121), that in the evolution of fundamental physical theories 'the beauty sometimes survives when the principles themselves do not'.

Careful serial reasoning has an important place in science. It can cover large parts of the logical ground very efficiently, especially when used in its more abstract mathematical forms, handling vast numbers of possibilities at once. In some cases it can help to expose the aesthetic beauty and economy, and the insightfulness, of the theoretical structures that physicists like Dirac, Einstein, Feynman, Salam, Weinberg, and others have used to such awesome effect. Careful serial reasoning systematises and economises large parts of the consistency checking process, by eliminating needless circularity. Knowing so and so's theorem might save you 20 years' hard labour. One needs to know what is serial and what is circular, and how big the circles are. What gets us into trouble is the notion that serial reasoning is *the* way to think – the notion that it is the *only* logical tool – especially as such a notion makes the impossible demand that all definitions, axioms, and deduction rules must be finalised before any reasoning can take place. Taken literally, this would stop us from discovering anything new about the world, because the thing would have to be defined before it could be discovered.

Mathematics, music, and the unconscious power of abstraction

I want to return briefly to mathematics, music, and the origin of the Platonic, hinted at in Part I. The model fitting hypothesis implies that perceptual processing handles – somehow – a combinatorially large number of possibilities. Such processing must therefore have a way of handling vast numbers of possibilities at once. This requires, or one could say amounts to, an unconscious power of abstraction.

The perceptual grouping of dots in Fig. 3 illustrates the point. Relative spatial proximity, 'motion as one', and 'belonging to a group' are abstract properties. They are abstract in the sense of being general enough to include or exclude vast numbers of special cases, hence vast numbers of model fitting possibilities. The same goes for the perceptual grouping of similar or related sounds, basic to speech and music perception,[80,103] as with the groups of notes evoking harmony changes in the musical examples of Figs. 2 and 4.

Perceptual grouping is recognisably a Platonic seed – part of the unconscious, primeval origin – of the consciously abstract ideas of set theory and integer used in mathematics. These ideas include for instance a recognition that the sequence of positive integers 1,2,3,... has no end. One can think of adding 1 to any positive integer, arbitrarily chosen, or equivalently of adding one object to any group of objects arbitrarily chosen. Though simple, the example is typical of mathematical thinking. I take this a little further in the Appendix.

Again, consider structure-from-motion perception, further generalising 'motion as one'. The walking lights demonstration is a sufficient example. It shows among other things that our visual system, after normal development in infancy, has unconscious knowledge of the branch of mathematics called three-dimensional Euclidean geometry.[77] Images like the Necker cube (Fig. 1 above), or any perspective drawing, show the same thing. In this sense the conscious discovery of Euclidean geometry by human mathematicians was a Platonic discovery, a discovery of something 'already there', transcending the concrete. To put it the other way round, infants groping for and looking for nearby objects are, among other things, unconsciously learning Euclidean geometry. They are abstracting geometric principles from a small number of concrete examples, in other words building general purpose internal model components that incorporate the relevant visual abstractions and idealisations, to which mathematicians, in later life, attach words like 'angle', 'straight line', 'plane', 'projection', 'perfect circle', 'rotation', 'translation', 'dilation', 'optical flow', etc. Someone deprived of this unconscious early learning cannot, in later life, cross the road safely, let alone – imagine it! – land an aircraft.[92,117]

And again, consider the phenomenon of illusory contours, illustrated in Fig. 5 overleaf. Their geometrical properties, including the strikingly smooth shape, are abstract properties – general enough to include vast numbers of special cases of smooth object outlines, from the edge of the moon and the edge of the sea to the curve of a perfect banana or elephant's tusk, the arch of a leaping dolphin, or the outline of a hanging grape or drop of water. Here is a Platonic seed for the conscious abstract ideas that mathematicians call curvature, continuously turning tangent, rate of change, and extremum principle, part of what is called differential geometry and the calculus of variations in space and, by extension, in spacetime.[118]

Indeed, we directly experience the extension to spacetime. We feel it kinaesthetically, through our sense of continuously evolving motion, for instance through athletic grace in humans and in other creatures and through the subtlety and elasticity of dance and music. In Fig. 5 the shape and clarity of the illusory contour is sensitive to the precise locations of the ends of the black segments. In a musical performance the kinaesthetic feel, the motion and continuity, is similarly sensitive to the precise timings of the onsets of sounds. The discussion of acausality illusions suggested that 'precise' means milliseconds. Even so gifted a musician as Mozart had to earn, by hard work at the keyboard, his legendary ability to make fortepiano passages sound like 'flowing oil'.[119] One might say that the difference between the performances of a great pianist and an average pianist is, in this respect, like the difference between the inner and outer illusory contours in Fig. 5.

Musical performances that successfully use rhythmic elasticity or rubato have a kinaesthetic feel that can be likened, in the spacetime analogy, to illusory contours that bend in interesting ways without breaking up. Such performances can have analysis-defying qualities of aliveness, as with organically changing images of living creatures in graceful motion, which we inwardly imitate. If our cave painting ancestors could have made movies, they surely would have. And none of this is accidental. As already noted, there are compelling biological reasons why perceptual processing has to treat space and time together. The spatiotemporality of internal models is a matter of life and death.

Constructivism versus Platonism: a false dichotomy

The hypothesis that perception works by model fitting points toward reconciliation of the constructivist and Platonic views both of mathematics and of music, indeed toward a natural harmony between those views.

The constructivist view is right, indeed far more so than is often realised. But a great deal of the construction involved is automatic and beyond conscious reach – including the construction of a vast repertoire of model subassemblies or 'schemas'[80] in infancy, in childhood, and beyond.

The Platonic view is right, because much of what mathematicians and composers and other artists feel they discover rather than construct, and find unutterably marvellous, is something 'already there', at least potentially, within the vast resources of the subconscious brain, our heritage from hundreds or thousands of millennia of biological evolution (explaining why mathematics, in particular, has real-world relevance[114]). What is 'already there' is a manifestation not only of our prodigious, and unconscious, early learning in the womb, in infancy, and in young childhood, but also of the subtle, multifarious, biologically ancient genetic inheritance that enables and drives such learning. The illusory contour in Fig. 5 is constructed, unconsciously. To wonder at its perfection is to glimpse the Platonic.

What is 'already there' includes the unconscious power of abstraction itself, and the unconscious interest in coherence and self consistency, on both of which the model fitting process we call perception depends and without which we could not handle – could not begin to cope with – the combinatorially large tree of possibilities that confronts us every millisecond of our lives. To think that the constructive excludes the Platonic, and vice versa, is yet another of our false dichotomies. The evidence is clear that both views articulate complementary, and indispensable, aspects not only of music and of mathematics,[120] but of perception and cognition itself and our very existence and survival.

The Platonic has its dark side. This is relevant, among many other things, to the crises in science policy and public understanding and to the threatened crisis in democracy itself. What is fundamental here is faced up to in Part III.

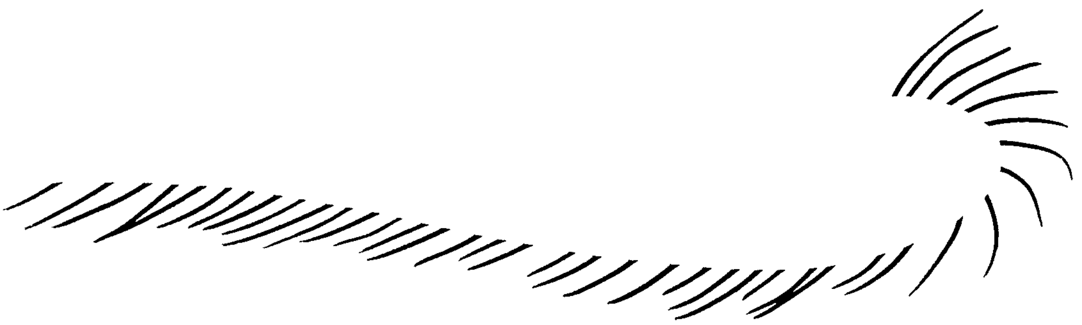
Appendix

On mathematical truth

In a remarkable recent book,[96] the mathematical physicist Roger Penrose presents speculations about brain function and consciousness that rely heavily on an intuitive view of what human 'understanding' is, in particular mathematical understanding – the vivid and compelling 'perception of unassailable mathematical truth' experienced by mathematicians, palpably independent of, even if built on, particular mathematical techniques and arguments. I share Penrose's sense of wonder at this, but I also suspect that, when it comes down to biological fundamentals, mathematical perception is not different in kind from other forms of perception and cognition. I would argue that it is profoundly and fundamentally similar, for the reasons hinted at in the main text. Indeed it could hardly be otherwise if you accept that we are part of the biosphere, and that we evolved from simpler organisms. This has implications for the arguments in Ref. 96 about algorithmic soundness and brain function, as I shall now try to show.

The model fitting hypothesis for perception suggests that vivid and compelling perceptions are associated with superlative goodness of fit, as judged by the brain's perceptual–cognitive apparatus. My visual examples included the illusory contour in Fig. 5 and the smoothly curved outlines of the moon's disc and the 'edge of the sea'. Let us imagine the edge seen under the clearest of skies, as I sometimes used to see it in my childhood in New Zealand, in the marvellously clear air of the Southern Hemisphere: an edge of the most exquisite sharpness, exactness, and perfection; one would naturally say a 'mathematical' curve. One would naturally say the same thing of the curved outline of a hanging drop of water, seen in perfect focus through a magnifying lens.

Now the brain's unconscious interest in coherence and self consistency of internal models, and in goodness of fit to sensory data, must be presumed – for the reasons most cogently argued, to my knowledge, in Ref. 85 – to extend to goodness of fit *between* different internal model structures, i.e. to extend to what we call cognitive association. Subjectively, math-



5 Simple demonstration of an illusory contour, a white or whitish edge grazing the inner ends of the black marks. Examples with still sparser data, and still showing a wonderfully smooth curvature, have been much studied and can be seen for instance in Fig. 2–6 of Marr's book[77] and Fig. 15 of Crick's book.[86] In constructing the contour, which does not exist physically on the paper, the visual system is unconsciously solving what mathematicians call an extremum problem of the calculus of variations

ematical perception seems to me to be an experience of sharpness, exactness, and perfection in certain kinds of cognitive association, which involve our unconscious power of abstraction in a special way. Indeed, mathematical experience expands and extends the power of abstraction, by small steps and occasional leaps, conscious and unconscious, into a world seemingly far removed from everyday experience. This is a combinatorially vast inner world, indeed an infinite world by extension, the Platonic mathematical world. Those who make journeys into it have encountered, again and again, self consistent structures of supreme beauty that fit together, as we perceive them, sharply, exactly, perfectly, and intricately. This is all the more wonderful because some of the structures that thus fit together seem quite unrelated at earlier stages in the journey. This is what mathematicians call 'non-triviality' and 'depth'. A celebrated example is that of functions of a complex variable and their relation to the prime or indivisible numbers.[121]

But most mathematicians would admit, I think, that mathematical perception begins with – grows organically from – perceptions of a much simpler kind, as already hinted in the main text. They would admit, I think, that the whole idea of 'unassailable mathematical truth' begins at the beginning of the same journey, with the simplest things like noticing the difference between two apples and three apples, and then, as already mentioned, noticing and making conscious the underlying abstraction – what we 'always knew about' intuitively – that in principle there is no end to the process of adding one more apple to the pile, or one more dot to the dot pattern, that there is no end to the sequence of integers 1,2,3,..., that the sequence is what mathematicians call 'infinite'. It seems to me that the perception of unassailable mathematical truth has exactly this character. We can perceive it, in this case the infinitude of the sequence of integers, through our unconscious power of abstraction – despite not having the slightest ability to imagine directly the size of integers like $10^{10^{10^{\cdots}}}$.

Going on from such beginnings, one may notice the endless sequences of dot patterns associated with the integers and their arithmetic: squares, triangles, rectangles, parallelograms, and other shapes made from rows of dots, their higher-dimensional counterparts, and the endless ways of dissecting them, as with 'cubes into hexagons' (Ref. 96, §2.4, p. 71). A few steps further on is Euclid's famous argument showing, unassailably as before, that the sequence of prime numbers is infinite (Ref. 96, §2.6/Q5, p. 80). Much further on – historically, many centuries on – we encounter the first really mindboggling non-triviality, the discovery that the dot patterns are intimately, organically, and precisely related to smooth curves like the illusory contour of Fig. 5, through what are called the theory of real numbers and the calculus of functions of real numbers. This has been known, and accepted as unassailable mathematical truth, for only a few centuries, with an increasing variety of cross-checks and progressively tightening standards of evidence for 'unassailability'. And to get from there to such things as the relation between prime numbers and complex functions required major expeditions in several further directions,[121] opening up routes still being explored today – to say nothing of the epic journey to Fermat's last theorem, only just completed.[122]

Reference 96 reminds us that, throughout such endeavours, mathematicians use a combination of intuitive perception and serial reasoning and that the latter is a mechanical, algorithmic process conforming to prespecified rules, hence programmable on a conventional electronic computer with a large enough memory, hence equivalent to what is called Turing computation. This is important for consistency checking and for the refinement and correction of intuitive perception – for keeping conscious and unconscious model fitting connected to each other. Crucial here is the notion of 'soundness' of an algorithmic process having a declared mathematical purpose, such as determining whether certain mathematical statements are true or false. Soundness means that the rules defining the algorithm are not only logically self consistent, but also such that the algorithm never delivers a result contradicting intuitively perceived unassailable mathematical truth (Ref. 96, §2.5). Thus an algorithm whose purpose is to determine whether there is a largest integer, or a largest prime number, would be called unsound if it ever came to a halt and gave the answer yes. There are of course less trivial examples.

Now the central point in Ref. 96 is that, whatever is going on in the brain of a human mathematician intuitively perceiving mathematical truth, for instance through the shapes and symmetries of dot patterns (Ref. 96, §2.4), it cannot be equivalent to a Turing computer running an algorithm that is sound in the mathematical sense just referred to. This is a consequence of Gödel's theorem, and is unassailable in the same way as the infiniteness of the integers is unassailable.

The deduction from Gödel's theorem is also consistent with the two hypotheses I am making here, first that mathematical perception is fundamentally similar to ordinary perception, and second that both work by unconscious model fitting. For those hypotheses imply that – however definite, and unassailable, the difference between two dots and three dots may be, and however vividly exact the edge of the sea or of a hanging drop may seem to be – perception, when viewed algorithmically, at its most basic functional level, in the way it uses neurons and protein molecules, must be an inherently approximate process. Indeed, if one thinks of perceptual processing in algorithmic terms, then the algorithms involved must have purposes far removed from that of determining mathematical truth, even supposing that the notion of 'purpose' has any meaning in this context. Such algorithmic purposes, even if meaningfully definable, would have to include things like measuring approximate goodness of fit, and pruning combinatorially large trees of possibilities. Moreover, as a biological imperative, such purposes would have to be achieved quickly and efficiently rather than perfectly and infallibly. Infinite accuracy and zero error rate have zero biological importance, whereas speed and efficiency have overwhelming biological importance. This predicts that algorithmic soundness – for any purpose at all – is unlikely to be relevant to a biological system like the brain at its most basic functional level; and this is enough for consistency with the deduction from Gödel's theorem.

It is noteworthy, in this connection, both that recent developments in electronic computing have recognised the power and practical usefulness of unsound but highly efficient algorithms,[123] and that the theory of computation implies that an algorithm may well, in any case, be 'its own shortest description' in a large number of instances, not least in biological systems (e.g. Ref. 87, p. 22). Thus the notion of 'purpose' need not enter the picture at all.

The foregoing, then, provides an alternative to Ref. 96's suggestion that the action of the brain must be non-Turing and must therefore use a novel kind of quantum computation. To be sure, the suggestion of quantum computation has a certain seductiveness in itself, because quantum computation has vast potential for solving combinatorially large problems,[116,124] which the brain needs to do anyway, whether for consciously mathematical or other reasons. Whether a role for quantum computation would *ipso facto* imply failure of Turing computability is yet another question, which might well be decided, in the end, on purely physical and not biological grounds.[96,114,125] I shall resist the temptation to speculate further, except to say that my own bet, for the time being, is what molecular biology suggests: that the brain is equivalent to a Turing computer of an unimaginably complex kind, running algorithms that are unlikely to be sound with respect to any independently definable purpose – essentially the kind of picture described in Refs. 85 and 87.

Another remarkable recent book,[116] by another mathematical physicist, David Deutsch, is of interest here; it presents a view very different from that of Ref. 96, closer to that of Ref. 114 on the matters under discussion and complementary to my own view. Chapter 10 of Ref. 116 on the nature and history of mathematics asks why outside-world reality should be perceivable and understandable at all, even to a limited extent, and points out the intimate connection between understandability on the one hand, and self consistency and abstractability thence, at least implicitly, mathematics and computation on the other. Mathematics is, on this view, not merely an ingeniously constructed aid to understanding physical reality, but intimately an aspect of physical reality. The detailed arguments elaborate the point I made in Part I, that '... self consistency of the internal model must mean, almost by definition, that the model is able or potentially able to represent something in the outside world'. The arguments in Chapter 10 of Ref. 116 imply that, even if we were unaware of examples as clearcut as those given above, we should still expect our perceptual–cognitive apparatus to have a significant unconscious knowledge of mathematics.

What emerges from all this, it seems to me, is that our intuitive understanding of mathematical truth stems not at all from the nuts-and-bolts *algorithmic activity* required to build and fit internal models but, rather, from the *model properties* themselves. Those model properties evolved under selective pressures to enable our ancestors to survive in the world, and therefore embody both abstractness and self consistency. Mathematics could be regarded as a way of consciously accessing, and stimulating the further development of, such model properties, many of which are normally beyond conscious reach.

Acknowledgements

I am grateful to the people and organisations listed in the Acknowledgements of Part I, especially to Max Perutz in connection with Note 88 below. John Coates kindly advised me on Note 122.

Notes and literature cited

The numbering continues from that in Part I.

72. M. E. MCINTYRE: 'Lucidity and science. I: Writing skills and the pattern perception hypothesis', *Interdisc. Sci. Rev.*, 1997, **22**, 199–216.
73. A. FRANKLIN: 'Are there really electrons? Experiment and reality', *Phys. Today*, 1997, **50**, (10), 26–33. One of the many centenary celebrations, and historical discussions, of what may reasonably be called the discovery of the electron.
74. J. DAVIES: 'Bacteria on the rampage', *Nature*, 1996, **383**, 219–220. News report on MRSA (methicillin-resistant staphylococcus aureus) and other emerging 'hospital superbugs' showing resistance to increasing numbers of antibiotics. Staphylococcal 'plasmids' are genetic packages of 'great architectural complexity' capable of being passed to other bacteria, and containing permutable components, perhaps like those of mammalian immune systems,[89] acting as 'free open reading frames (cassettes)', or, in the language of this essay, versatile model fitters. Such capabilities make the development and spread of resistance to any antibiotic almost inevitable, and underline the irresponsibility of the profligate use of antibiotics. Davies reminds us that 'The breakdown of public health measures due to famine, poverty, and war contributes to the dissemination of resistance, ... compounded by the growth of international travel ...'
75. S. H. SCHNEIDER: 'Laboratory earth: the planetary gamble we can't afford to lose'; 1996, London, Weidenfeld and Nicolson/New York, HarperCollins (Science Masters series), 184 pp. This gives a good lay person's explanation of some of the reasons for the large uncertainty about future environmental change – including what is simplistically called 'climate change' – and about what its early warning signals might be. For instance, there are straightforward uncertainties about the numerical magnitudes of critical parameters, concerning cloud droplet sizes for instance, and the rates and modes of transport of water, carbon dioxide, and other chemicals and pollutants in the atmosphere–ocean–cryosphere–biosphere system and the concomitant biological adaptations (e.g. *Pfiesteria piscicida*). Another, more subtle, reason is cogently argued in Ref. 76. Some of the oft-repeated claims, implicit or explicit, to 'infallible prior knowledge' of what's ahead, despite the large uncertainties – and despite the possibility of continuous, unstoppable sea-level rise for two or more centuries – are discussed in Schneider's book. More of them are collected and discussed in R. GELBSPAN: 'The heat is on: the high stakes battle over earth's threatened climate'; 1997, New York, Addison Wesley, 278 pp. For a sufficient example of an implicit claim to prior knowledge, *see* p. 46.
76. T. N. PALMER: 'A nonlinear dynamical perspective on climate prediction', *J. Clim.*, submitted Sept. 1997. The atmosphere–ocean–cryosphere–biosphere system is a nonlinear dynamical system. What is already known about the general behaviour of such systems[114] implies still greater uncertainty than one might gather from Ref. 75, especially about early warning signals. Palmer's paper uses simple but apt analogies to bring out these points.
77. D. C. MARR: 'Vision: a computational investigation into the human representation and processing of visual information'; 1982, San Francisco, Freeman, 397 pp. *See* also Note 5 of Part I.
78. M. HUNT: 'The story of psychology'; 1993, New York, Doubleday, 762 pp. More detail in Note 17 of Part I.
79. The Internet web and ftp sites are http://www.atmos-dynamics.damtp.cam.ac.uk/ and ftp://ftp.damtp.cam.ac.uk/pub/papers/mem/. All relevant file names begin with the eight characters 'lucidity'. They include a fuller version of the text with fuller references (lucidity.ps) together with relevant visual animations and audio[108] (PC .wav) recordings, for download in binary mode. The visual animations include two walking lights demonstrations, in MPEG files named lucidity-walking-lights1.mpg and lucidity-walking-lights2.mpg, and a rotating Necker cube demonstration in lucidity-necker.mpg. *See* also Notes 8 and 31 of Part I.
80. A. S. BREGMAN: 'Auditory scene analysis: the perceptual organization of sound'; 1990, Cambridge, MA, MIT Press, 773 pp. Further notes under Note 22 in Part I; on grouping in music, *see* also Refs. 6 and 23 of Part I.
81. J. DRIVER: 'Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading', *Nature*, 1996, **381**, 66–68. Visual cues, in the form of the moving lips of a human speaker seen on a television screen, cause a 'ventriloquist effect', an illusory displacement in the perceived spatial location of the corresponding sound source, like the illusory sound-source displacement that occurs when watching television in the ordinary way. The experiments show that this illusory displacement can help the listener to distinguish two speech signals emerging from a single loudspeaker. Thus the internal model assumes two spatially separated sound sources, leading to improved discrimination even though this worsens the fit to the actual acoustic data, there being only one sound source in the outside world. Notice incidentally that this is a non-trivial consistency check on the model fitting hypothesis itself; the results make sense only if a single internal model is being fitted to both sets of sensory data, visual and auditory. *See* also the excellent discussions in Bregman's book[80] of related phenomena like the 'duplex perception of speech' and the 'McGurk effect'; and *see* also Refs. 82–84 and 92. All these studies clearly suggest not only that perception works by model fitting, but also that different sets of sensory data – for instance auditory and visual – are fitted to the same internal model. Also Note 40 of Part I.
82. R. SEKULER, A. B. SEKULER, and R. LAU: 'Sound alters visual motion perception', *Nature*, 1997, **385**, 308. Another multisensory psychophysical experiment, with a clearcut result that is easy to make sense of if we assume that visual and auditory data are fitted to the same internal model. Experiments were controlled against a popular alternative explanation in terms of attention switching.
83. N. K. LOGOTHETIS, D. A. LEOPOLD, and D. L. SHEINBERG: 'What is rivalling during binocular rivalry?', *Nature*, 1996, **380**, 621–624; *see* also p. 587. Here 'binocular' replaces 'multisensory', but with the same implications. These new experiments add powerfully to the evidence for model fitting by showing that the percept, i.e. the internal model, chosen by the visual system is insensitive to changes in the route taken by the incoming sensory data. Included are experiments that

present conflicting stimuli to the two eyes and, at intervals of one-third of a second, swap which stimulus goes into which eye. The resulting pairs of percepts exhibit bistability not at intervals of one-third of a second but at longer, randomly variable time intervals, qualitatively like Necker cube flipping. This firmly rules out hypotheses that assume 1–1 mapping, like monocular dominance or fatigue.

84. G. PELLEGRINO, E. LÁDAVAS, and A. SARNÉ: 'Seeing where your hands are', *Nature*, 1997, **388**, 730. The results of this experiment, on a perceptual phenomenon called 'cross-modal extinction' resulting from certain kinds of brain damage, become intelligible if we hypothesise that the subject's brain was fitting two modes of sensory data, visual and tactile in this case, to a single internal model of his own hands and their spatial positions.

85. D. HOFSTADTER and co-workers: 'Fluid concepts and creative analogies: computer models of the fundamental mechanisms of thought'; 1995, New York, BasicBooks, 518 pp.

86. F. CRICK: 'The astonishing hypothesis'; 1994. London, New York, Simon and Schuster, 317 pp. *See* caveat in Note 32 of Part I.

87. S. KAUFFMAN: 'At home in the universe: the search for laws of self-organization and complexity'; 1995, New York, London, Viking Penguin, 321 pp. Also Notes 1–4 of Part I.

88. I am grateful to Max Perutz for kindly vetting my statements about the immune system. He prefers to call the immune system's internal models 'templates' – because the model information resides in the shapes of 'antibody' molecules that complement or fit the incoming material hence 'recognise' it and clamp on to it – and to say that the system works by, in his words, 'constructing, by a stochastic mechanism, many millions of different templates, one or other of which may happen to be complementary to any particular chemical compound or viral or bacterial component'. Once the particular compound or component is thus recognised, 'the cell carrying that template on its surface is stimulated to proliferate and make multiple copies of that template' (M. F. PERUTZ: personal communication). Such proliferation, of a template pattern selected, in a very Darwinian sense, by its goodness of fit to the incoming material, is the main mechanism whereby the immune system can be said to 'fit an internal model' to the incoming data. The system's repertoire of possible models – its 'many millions of different templates' – is constructed stochastically, by a kind of 'genetic roulette', from pre-existing, permutable model components or, so to speak, permutable building blocks, specified by permutable DNA segments. *See* Notes 89 and 126 for more detail, including recent work on how the system iteratively improves the goodness of fit.

89. *See* for instance T. FRÄNGSMYR and J. LINDSTEN (eds.): 'Nobel lectures in physiology or medicine'; 1993, Singapore, World Scientific, 584 pp, concerning the Nobel Lectures of Nils Jerne (pp. 203–225) and Susumu Tonegawa (pp. 373–405), delivered 8 December 1984 and 1987 respectively. This work is now seen as basic to understanding the more recently evolved parts of mammalian immune systems that give rise to what is called 'acquired immunity',[126] the system's perception and knowledge of detailed patterns in foreign material. Tonegawa's work, in particular, led to molecular-level knowledge of the actual mechanisms whereby the immune system plays 'genetic roulette' to construct, by trial and error, its basic repertoire of approximate internal models available to be fitted to patterns in incoming foreign material.[88] Recent progress in understanding how the system iteratively improves the goodness of fit is reported in *Science*, 1997, **276**, 1658–1659 and references therein, particularly G. J. WEDEMAYER, P. A. PATTEN, L. H. WANG, P. G. SCHULTZ, and R. C. STEVENS: 'Structural insights into the evolution of an antibody combining site', *Science*, 1997, **276**, 1665–1669. An ingenious experiment has illustrated in great detail, in one case, just how far the fit can be tightened through 'the well understood process of antibody maturation'.

90. C. SAGAN: 'The demon-haunted world: science as a candle in the dark'; 1996, London, Hodder Headline/ Random House, 436 pp. Chapter 6 gives a clear summary of the evidence concerning the perceptual phenomenon called hallucination. *See* also Refs. 91 and 92.

91. C. D. FRITH and P. FLETCHER: 'Voices from nowhere', *Critical Quarterly*, 1995, **37**, (2), 72–83. This brief review, written by a research neuropsychologist and a research psychiatrist, gives a readable and well-referenced discussion of the auditory hallucinations reported by people suffering from schizophrenia. These occur in 'about 65 per cent' of cases, and are experienced as wholly real, often distressingly so. *See* also Ref. 111.

92. O. SACKS: 'An anthropologist on Mars: seven paradoxical tales'; 1995, New York, Alfred Knopf, 330 pp. *See* also Note 38 of Part I, including remarks about the permanent adaptive visual hallucination of the patient 'Greg F', who was blind yet 'watched television'.

93. Again *see* the later section 'On epistemology' etc.; and to get a more detailed idea of the 'difficulties in the philosophy of science' *see* for instance A. F. CHALMERS: 'What is this thing called science? An assessment of the nature and status of science and its methods', 2nd edn; 1982, Milton Keynes, Open University Press, 179 pp. This summarises recent thinking among historians and philosophers of science, carrying on from naive inductivism to Popper, Lakatos, Kuhn, and Feyerabend. In Chalmers' own admirably honest words, 'We start off confused and end up confused on a higher level' (p. xix).

94. *See* the historical remarks on pp. 18–29 of Ref. 92, also the chapter 'To see and not see', about a personal tragedy that could have been avoided, almost certainly, had those involved had the slightest inkling of how perception works. For another such case, *see* Ref. 117.

95. E.g. J. L. MILLER and A. M. LIBERMAN: 'Some effects of later occurring information on the perception of stop consonant and semivowel', *Perception Psychophys.*, 1979, **25**, 457–465. Here the acausality illusion is referred to as an 'after going effect'. Other terms, such as 'backward referral in time', can be found in the literature. I have not yet come across the simpler and more straightforward term 'acausality illusion'.

96. R. PENROSE: 'Shadows of the mind: a search for the missing science of consciousness'; 1994, Oxford, Oxford University Press, 457 pp. Also Note 53 of Part I, and the Appendix to Part II, the present article.

97. Ref. 96, pp. 385–387, gives a brief discussion of the neurological experiments of Benjamin Libet and the controversies they have triggered; more extensive discussions and debates, with large bibliographies, can be traced through Refs. 99, 100, and 111. There have even been suggestions to the effect that physical causality principles be abandoned.

98. N. HUMPHREY: 'A history of the mind'; 1992, London, Chatto & Windus, Vintage, 230 pp. An important reference on the nature of perception and consciousness. *See* also Note 33 of Part I.

99. D. C. DENNETT: 'Consciousness explained'; 1991, London, Penguin, 511 pp. This insightful book by a well known philosopher discusses several examples of what I am calling acausality illusions, including the important Grey Walter slide projector experiment, apparently not published elsewhere. On the consciousness debate itself, *see* also Ref. 98, and D. C. DENNETT: 'Kinds of minds: towards an understanding of consciousness'; 1996, London, Weidenfeld and Nicolson New York, HarperCollins (Science Masters series), 184 pp.

100. D. C. DENNETT AND M. KINSBOURNE: 'Time and the observer: the where and when of consciousness in the brain', *Behav. Brain Sci.*, 1992, **15**, 183–247. *See* also correspondence in *Behav. Brain Sci.*, 1995, **18**, 810–811.

101. An adequate review of the literature on cortical processing timespans is beyond our scope here, but it can safely be assumed that the relevant timespans for multi-level processing are far longer than a few milliseconds, and likely to be of the order of hundreds of milliseconds. Direct electrophysiological evidence, from cortical activity evoked by sudden perceptual stimuli, visual or auditory, is illustrated by the ensemble averaged, event related electroencephalograms shown in Fig. 35 of Crick's book,[86] Chap. 9, p. 112. It is typical for the activity observed in this way to continue for '400 msec or longer' (C. D. FRITH: personal communication). In Crick's examples, the ensemble averaged signals last around half a second, each showing one peak near 200 msec and another near 300–400 msec. Crick states that a prominent peak near 300–400 ms is 'fairly common', and 'usually correlates with something that is surprising', or attention grabbing or eye catching, like visual 'pop out' (p. 111 and caption to Fig. 35). Psycholinguistic experiments reveal similar timespans, corresponding to two or three syllables of speech, for pre-conscious processes in speech perception. Examples include the work of Swinney, Seidenberg, Tanenhaus, and others summarised in Ref. 104, p. 143 and in pp. 210–211 of S. PINKER: 'The language instinct: the new science of language and the mind'; 1994, London, Allen Lane (Penguin), 494 pp. The same goes for music perception, as in the simple examples described in the text.

102. Anyone with normal hearing can easily and effortlessly tell that the sound of the piano note precedes the sound of the pip in such cases. Audio demonstrations are available on the Internet[79] in the PC .wav files lucidity-radiopips100.wav and lucidity-radiopips200 .wav, giving the 100 and 200 ms cases respectively (low quality, each just over 100 kbyte), and in the MIDI file lucidity-radiopips.midi (just over 200 byte), giving both cases together. In musical language, the effects in the 200 and 100 ms cases could be described as a leisurely 'dotted rhythm' and 'grace note' respectively.

103. S. HAWKINS: 'Arguments for a nonsegmental view of speech perception', in Proc. Int. Cong. on 'Phonetic science', Symp. on 'Dynamic, nonsegmental approaches to phonetics', (ed. K. Elenius and P. Branderud), Vol. 3, 18–25; 1995, Stockholm, University of Stockholm and Royal Institute of Technology (KTH), ISBN 91 7170 836 7.

104. R. JACKENDOFF: 'Languages of the mind: essays on mental representation'; 1992, Cambridge, MA, MIT Press, 200 pp., ISBN 0 262 10047 9. *See* the allusion to 'temporal anomaly' on p. 141, and the hypothesis of 'parallel multiple analysis' on pp. 140–145, the latter first published in *Music Perception*, 1991, **9**, 199–229.

105. A crude realisation of this demonstration of apparent motion (with inadequate control over timing, because timing depends on the browser software or other system used for viewing) is available on the Internet[79] in the MPEG file lucidity-apparent-motion.mpg.

106. We also tried the bicolour version, a red disc followed by a green disc, discussed by Dennett[99] under the heading 'color phi',[78] in order to test his interesting claim that a colour change from red to green is perceived at a definite intermediate time. However, we could not get this to work convincingly, even though we used the exact timings mentioned in Dennett's book,[99] namely, 150 ms on and 50 ms off (cf. 200 ms on and 200 ms off, as used in the demonstration described in the text). We clearly saw a red disc departing and a green disc arriving, but no definite perceived time for the colour change.

107. U. LE GUIN: 'The dispossessed', Chap. 6; 1974, London, Gollancz/1996, London, HarperCollins, 319 pp.

108. Recordings of the musical examples used in the text are available on the Internet,[79] for download in binary mode, as audio PC (.wav) files at sampling rates 22 and 44 kHz (standard compact disc quality). The recordings use a standard high quality computer driven sound module. Though unable to match the delicacy and subtlety of the best human performances, they are accurate realisations of the written scores and sufficient illustrations of the points under discussion. The files, in the order mentioned in the text, are named lucidity-piano1..., lucidity-orch1..., lucidity-orch2... (the 1–1 mapping version), lucidity-piano2..., and lucidity-orch3... (the last two corresponding to Fig. 4), where the dots denote a suffix like -22m.wav or -44s.wav indicating the sample rate and whether mono or stereo, -22m.wav files being the smallest, about 0·7 Mbyte each. The corresponding written scores are in the graphics files lucidity-piano12.gif, lucidity-orch1.gif, lucidity-orch2.gif, and lucidity-orch3.gif. I am grateful to Ben Finn of Sibelius Software (http://www.sibelius-software.com/) and Jeffrey Ginn of Ginn Music (http://www.ginn-mus.demon.co.uk/) for lending their professional expertise with the audio recordings, and to the composers Alexander Goehr, Robin Holloway, Virginia Seay Ploeser, Yuval Shay-El, Roderick Skeaping, and Hugh Wood for advice on the scoring. (They all agree that an accompaniment designed to preserve the musical sense of the piano solos would have to change harmony at, not after, the time of the arrow.)

109. T. G. BEVER, M. F. GARRETT, and R. HURTIG: 'The interaction of perceptual processes and ambiguous sentences', *Memory and Cognition*, 1973, **1**, 277–286.

Easier now from www.atm.damtp.cam.ac.uk/people/mem/papers/LHCE/

110. S. JONES: 'In the blood'; 1996, London, HarperCollins, 302pp. A witty and insightful extended essay, by a respected geneticist, on what we know and do not know about human genetics. Jones takes good care to debunk the simplistic racist, eugenics, nature–nurture and genetic engineering myths. On these points *see* also, for instance, among many others, Wills,[127] P. MEDAWAR: 'Science and the sanctity of life', in 'Pluto's republic', 311–323; 1982, Oxford, Oxford University Press, and K. E. DAVIES, A. J. CLARKE, and P. S. HARPER: 'The genetic revolution and medicine in the 21st century', *Europ. Rev.*, 1997, **5**, 39–54 ('Is having the wrong genes going to become an acceptable line of defence in a court of law in cases of violence? Fortunately, ... the biological basis of behaviour ... is unimaginably complex, and is open to modification by numerous environmental influences.').

111. S. A. SPENCE:'Free will in the light of neuropsychiatry', *Philos. Psychiat. Psychol.*, 1996, **3**, 75–90; *see* also pp. 91–100 for commentaries by C. Frith, B. Libet, G. L. Stephens, and reply by Spence.

112. J. COHEN and I. STEWART: 'The collapse of chaos: discovering simplicity in a complex world'; 1994, New York and London, Penguin, 495 pp. This gives an excellent discussion of the intimacy of nature–nurture interactions, 'a rich, fascinating, and largely unexplored joint dynamic', p. 314ff., convincingly arguing, like Wills,[127] Kauffman,[87] Jones, and others,[110] that the genes are only part of an immensely more complicated story. *See* for instance the section on 'genetic assimilation', otherwise known as the 'Baldwin effect'.[127]

113. H. BONDI: 'Assumption and myth in physical theory'; 1967, Cambridge, Cambridge University Press, 88 pp.

114. I. STEWART: 'Does God play dice?', 2nd edn; 1997, London, Penguin, 401 pp. Along with Ref. 112, this gives an insightful view of how science works as a model fitting process, and why mathematics is relevant: 'To criticise mathematics for its abstraction is to miss the point entirely' (p. 363). My remarks in the section 'On epistemology' in the text and in the Appendix are, in effect, an elaboration of this point.

115. S. WEINBERG: 'Dreams of a Final Theory – the search for the fundamental laws of nature'; 1993, London, Vintage Books, 260 pp.

116. D. DEUTSCH: 'The fabric of reality'; 1997, London, Allen Lane, 390 pp.

117. R. L. GREGORY and J. G. WALLACE: 'Recovery from early blindness: a case study', in 'Concepts and mechanisms of perception', (ed. R. L. Gregory); 1974, London, Duckworth, 669 pp. This is perhaps the most careful and detailed psychological case study (of the total of 20 or so known cases[92]) on what happens when infantile visual model-building is prevented, or atrophies – analogous to Sacks' case in 'To see and not see',[92] and similarly tragic. To pursue the analogy drawn in Notes 88, 89, and 126, one might say that a visual system not exposed to data is, apart from the enormous extra complexity, like an immune system not exposed to invading material.

118. The mathematician Leopold Kronecker is reputed to have said, 'God made the integers; all else is the work of man'. He must have been unaware of illusory contours.

119. Mozart admitted this himself; *see* H. MERSMANN and M. M. BOZMANN: 'Letters of Wolfgang Amadeus Mozart'; 1928, 1972, New York, Dover, 278 pp. Relevant quotes can be found, for instance, on pp. 40 and 221.

120. The creation of anything worth creating, in the arts and sciences equally, always seems to involve an intricate interplay between conscious and unconscious construction. This is well described on pp. 191–196 of Ref. 121. *See* also 'brain as a committee', p. 292 above. There are the celebrated 'eureka moments': 'illumination, which can happen in a fraction of a second, is the emergence of the creative idea into the conscious', and there is the less celebrated, but arduous, preparation for such moments, impossible without 'an intense conscious curiosity about the subject ... a craving to exercise the mind on it, quite like physical hunger', lasting for many years. There is the need to find ways, different for different individuals, of 'giving the subconscious every chance'; *see* also Note 34 of Part I. There is the 'devastating experience' of losing the curiosity and the drive to undertake such arduous labour. All this should be required reading for science policymakers and bureaucrats.

121. J. E. LITTLEWOOD: 'A mathematician's miscellany'; 1953, paperback reissue as 'Littlewood's miscellany', with further material (ed. B. Bollobás; 1986, Cambridge University Press, 200 pp). *See* pp. 5 and 89 for comments on the relations between prime numbers and complex functions. *See* also the vivid description, on pp. 249–256 of Ref. 96, of how complex numbers, a key stage in the journey toward complex functions, were discovered, against all the odds, by Gerolamo Cardano in the sixteenth century.

122. Fermat's last theorem, that the equation $x^n + y^n = z^n$ has no positive integer solutions (x, y, z) when n is an integer greater than 2, has challenged mathematicians ever since Pierre de Fermat stated it well over 300 years ago. The chapter on it in Ref. 121, pp. 74–79, sketches a view from the 1920s; the theorem was not proven until the 1990s. The proof draws on deep and wide-ranging concepts, far beyond anything Fermat could have known, and was discovered by the ex-Cambridge mathematicians Andrew Wiles and Richard Taylor. Readers interested in a brief commentary by an expert in the field may consult J. H. COATES: 'The work of Andrew Wiles', *Not. Am. Math. Soc.*, 1996, **43**, (7), 760–763. A more substantial survey for a nonspecialist but mathematically literate reader, explaining some of the key ideas but stopping short of the full, and very formidable, technicalities, is that of K. RIBET: 'Galois represention and modular forms', *Bull. Am. Math. Soc.*, 1995, **32**, 375–402. Full technical details are given in a 139-page survey of the problem and its history by H. DARMON, F. DIAMOND, and R. TAYLOR: 'Fermat's last theorem', in 'Elliptic curves, modular forms, and Fermat's last theorem', 2nd edn, (ed. J. H. Coates and S.-T. Yau), 2–140; 1997, International Press. I am grateful to John Coates for providing these references.

123. E.g. work on 'genetic algorithms' at the Santa Fe Institute and elsewhere; also e.g. B. A. HUBERMAN and T. H. LUKOSE: 'An economic approach to hard computational problems', *Science*, 1997, **275**, 51–54. 'Portfolio' or risk-spreading algorithms make tradeoffs between speed and algorithmic unsoundness.

124. P. W. SHOR: 'Algorithms for quantum computation: discrete logarithms and factoring', Proc. 35th Ann. Symp. on Foundations of Computer Science, 124–134; 1994, New York, IEEE Press.

125. T. N. PALMER: 'A local deterministic model of quantum spin measurement', *Proc. Roy. Soc. London*, 1995, A **451**, 585–608. The idea is to replace quantum non-locality by algorithmic non-computability arising from 'riddled basins' of attractors in phase space. For an excellent lay person's description *see* Ref. 114, 2nd edn, pp. 348–356.

126. D. T. FEARON: 'Seeking wisdom in innate immunity', *Nature*, 1997, **388**, 323–324. A useful discussion of acquired and innate immunity, summarising an emerging understanding of how the evolutionarily more ancient components of a mammalian immune system seem to direct the 'attention' of the less ancient model fitting process, as I am calling it. The more ancient components, 'innate' in the sense of being fixed in genetic memory, and lacking a large repertoire of models to fit to incoming patterns, are found also in invertebrates and seem more oriented to detecting general damage caused by invaders. In mammals, such damage or 'danger' information seems to be part of what tells the model fitting system[88,89] when and where to look for an invader, and part of what stops it looking for and attacking its owner's own tissues. *See* also, for example, earlier correspondence and controversy in *Science*, 1996, **272**, 1405–1408. Such an attention-directing role of more ancient parts of the system seems to me to deepen the partial analogy between immune system function and brain function.

127. C. WILLS: 'The runaway brain'; 1994, London, HarperCollins, 358 pp. *See* also Note 13 of Part I.

The manuscript was received 15 May 1996.

Professor Michael McIntyre
Centre for Atmospheric Science at the Department of Applied Mathematics and Theoretical Physics
Silver St
Cambridge CB3 9EW
UK
mem@damtp.cam.ac.uk
http://www.atmos-dynamics.cam.ac.uk/

Michael McIntyre is Professor of Atmospheric Dynamics and Co-director of the Centre for Atmospheric Science in the University of Cambridge. He is a fellow of the Royal Society, a member of the Academia Europaea, and a Rossby medallist of the American Meteorological Society. His main research is on theoretical fluid dynamics; he served for 10 years as an editor of the *Journal of Fluid Mechanics*, and can tell you about the fluid-dynamical mechanisms that allow northern pollutants to cause a southern ozone hole. He has also had a longstanding interest in perception and cognition. It arose partly from the problem of visualising atmospheric motion in ways that connect intuitive perceptions with mathematical theory, and partly from interests in musical performance, musical composition, and musical acoustics. His 1978 *Interdisciplinary Science Reviews* article with James Woodhouse on the acoustics of stringed musical instruments was recently republished by the Acoustical Society of America as part of a new benchmark collection on musical acoustics.