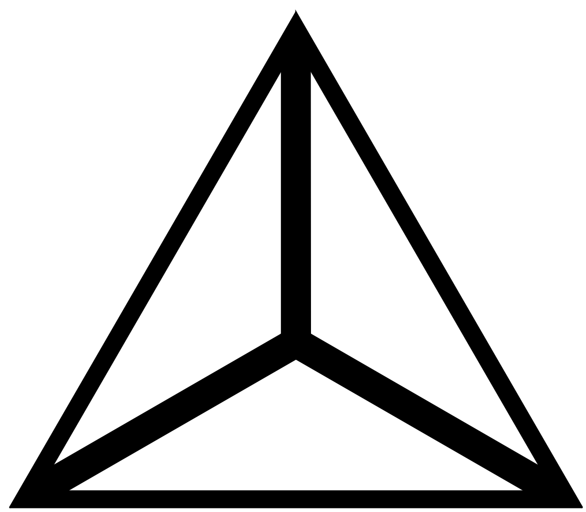
Visions of Mind and Body

a technological solution to the mind-body problem



Dr. James Anderson

Dedication

Its flavour when cooked is more exquisite far
Than mutton, or oysters, or eggs:
(Some think it keeps best in an ivory jar,
And some in mahogany kegs.)

The Hunting of the Snark page 81.

Hunting the Snark is my greatest adventure and Linda my greatest companion.

Preface

The author started his research career with his head in a cardboard box. This was not a particularly auspicious start, but it was a way to do experiments to discover how efficient human vision is. As well as the author's head, the box contained an oscilloscope, and a key pad to record how sure he was that he could see various patterns on the oscilloscope. It turned out that human vision is stunningly efficient.

After doing some mathematics to work out interesting patterns, and writing computer programs to analyse his experimental data, the author struck on the idea of writing programs that can see for themselves. At a stroke he took his head out of the cardboard box and has never looked back.

After a while the author married and, though he does not have any children, he does share a house with a very sweet, but deadly, cat. This set the author to thinking how intelligent a cat can be, what its feelings are, what it is conscious of, how it can have free will, and how it is that the cat, the author, and his wife can co-exist peacefully in one house. The answer to these questions, and many more, is set out here, but not in the way you might expect.

You might expect the answers in a book to be written in words, and so they are, but the words point to geometrical patterns, or *visions*, that explain how a robot can take the place of a cat. There are plenty of words and pictures in the body of this book, and no equations, so it is easy to read. Once you build up the visions in your mind's eye you will *see* what intelligence, consciousness, feeling, and free will are. Whether this power of imagination comes from an inbuilt mathematics or magic, the author cannot say.

After writing this book the author will start to build the robot. No human lives long enough to build a robot that is like a cat or ourselves, but if mankind carries on scientific research for generations we should, eventually, produce a robot that, like a cat, or like us, is sweet and not too deadly.

The first four chapters introduce the perspex and the things it can do. The perspex is a mathematical entity that can describe the shape and motions of bodies and the shape and thoughts of brains made from artificial neurons. This is interesting in itself, to have a mathematical explanation of *mind* and *body*, but these four chapters are just the foundations for everything that follows.

The fifth chapter describes how an android might learn language in a human society, and deals with various technical limitations that language imposes on a mind. These limitations do not apply to a mind that is capable of continuous thoughts, as would be the case if an android could be equipped with a perfect perspex machine. Even in the case of an imperfect, perspex machine that is no more powerful than a digital computer, the perspex machine still shows us how to exploit geometry to compute many of the perceptions and experiences we hold most dear.

This chapter provides a link to the remaining chapters that explain how an android might be programmed so that it can experience visual consciousness, free will, intelligence, and feeling. The chapter also shows how it might be possible to relate these phenomena to words, so that an android can talk to us about its experience of the world.

The antepenultimate chapter explains how an android can experience time. This chapter also proposes an experiment to test an extreme hypothesis about the nature of time. This view of time, along with all the earlier material, is taken up in the penultimate chapter on spirituality – both for androids and for us.

The ultimate chapter: reviews the material in the book; sets out a vision of future research; and explains the paradigm of manifestation, so that others may undertake similar work more easily.

Dr. James Anderson, B.Sc., Ph.D., MBCS
Lecturer in Computer Science
The University of Reading
England

j.anderson@reading.ac.uk
http://www.cs.reading.ac.uk/people/staff

Acknowledgements

There is a saying that well begun is half done. I hope this is true, because it has taken me half a working lifetime of contemplation to arrive at the hypothesis set out in this book. I will now turn to testing the hypothesis and will use the book as a reference.

I hope the book will be of interest to the general reader and of help to the postgraduate, computer science student looking for a research topic. Many professional scientists, research students, and undergraduates have helped me in the research leading up to this book. They are all properly and cordially acknowledged in the published papers, most of which are not cited here.

I would now like to thank those who helped me during the writing of this book.

Steve Maybank continued to be a source of information on mathematical aspects of computability and helped me track down some of the text books cited here.

Claire Whitehead proof read the manuscript and made many helpful suggestions. The resulting manuscript is the compromise that arose when the irresistible force of proof reading met the unmovable object of dyslexia.

Linda Anderson, my wife, was a constant help and comfort during the tedium of writing this book. It is the largest, single, linguistic exercise I have undertaken and is more than I thought I could bear. It would be so much easier if, instead of writing, I could just show you the pictures in my mind.

Gerald Mogglington-Moggs, our cat, provided relief from the needless stresses of the working day and inspired, by his actions, some of the thoughts expressed here. I wonder if it is possible to bite the heads off the bureaucrats and drag their mindlessly twitching bodies through the cat flap?

I absolve all those who have helped me of any blame for errors, omissions, or infelicities in the book. I am entirely responsible for its contents.

Contents

$\vec{x}$ $\vec{y}$

Figures

$\vec{x}$ $\vec{y}$

CHAPTER 1 *Introduction*

When I started the research that lead to this book, it was with a very simple aim. I wanted to explain how a robot might be given a sense of sight that is like ours. Before long, it became clear to me that our sense of sight is related to everything we are and do. We see the world and decide how to move in it. We see our friends and talk to them. We see a sunset and feel its beauty. I soon realised I would have to explain all this, but can these psychological things be explained in terms of vision? After all, blind people can do and feel almost everything that sighted people can, so what use is a visual explanation of the human condition? Then I realised there is something deeper, something inside vision, that explains how our minds work, and how we relate to the world, whether we are sighted or blind. At the heart of vision there is a single, physical explanation of mind and body. This book explains how to capture that physical phenomenon so that a robot can be given a mind like ours.

What is at the heart of vision? Mathematicians might say that geometry is, because it describes the shapes of things in the world and how they look in perspective. They might offer projective geometry to the sighted who see in perspective, and Euclidean geometry to the blind who feel the position of things just where they are. With only a little effort, mathematicians can offer both kinds of geometry in one job lot. The simplest kind of geometrical figure is called a *simplex.* The simplex can describe everything in Euclidean geometry, it can describe all of the geometry that blind people need, but it cannot describe perspective in any useful way. For that a perspective simplex, or *perspex* is needed. A perspex can provide all of the geometry that sighted people need and it can mimic a simplex perfectly so it is perfectly useful to the blind. The perspex can describe the shape and motion of any body in the universe, be it something as simple as a snooker ball, or as complex as a human body and brain.

But perhaps we should give mathematicians short shrift? After all, mathematics does not *do* anything, it just *is*. In particular, projective geometry cannot describe figures or solve problems by itself it is just a set of rules that a mathematician uses to help to do these things. Projective geometry does not explain how a mathematician's mind works it defines only what a mathematician's mind should achieve in various circumstances to explain the geometrical shape and motion of bodies. Incidentally, this is why it is hard to be a mathematician: it is easy enough to learn the rules, but it is harder to know what to do with them, and very much harder to know how to invent new rules and prove that they are correct. Projective geometry provides a theory of body, but not a theory of mind.

What is at the heart of vision? The computer scientist might say that the Turing machine is because this theoretical computer explains how all computations are done. In particular, it explains how the computations of projective geometry are done so that a human being, or a robot, can see the world and interact with it. We should give computer scientists short shrift too. After all, they claim that a theoretical computer *is* not anything, it just *does*. According to them, the Turing machine is an abstract thing that has no physical relationship with the world; it cannot interact with any body in the universe, but it does offer an explanation of how minds work. The Turing machine provides a theory of mind, but not a theory of body.

Suppose you wanted to explain how minds and bodies work. What would you do if mathematicians offered you a theory of body, but not of mind, and computer scientists offered you a theory of mind, but not of body? I did the obvious thing. I unified the two theories in a mathematical paper[1], then wrote this book to explain, in plain English, how a theoretical perspex machine can operate on perspexes to do everything that a Turing machine can do, and more. Thus, the perspex machine provides a theory of mind. Practical perspex machines are made out of the physical stuff of the universe, so they are bodies. In theory, the perspex machine applies to everything; it is everything in the universe – every mind and body. The perspex explains the human condition and, as a side effect, the bat condition, cat condition, Martian condition, and everything else, including the insane dreams of opium smokers, and the interactions of snooker balls. All this because the perspex links physics and computation by unifying projective geometry and the Turing machine.

What is at the heart of vision? The neurophysiologist might say that neurons are at the heart of vision because they detect light in the retina, interpret it, and cause us to interact with the world, talk to friends, and feel the beauty of a sunset. We can give neurophysiologists short shrift too. The perspex is a neuron. The physical constraints of linking geometry to computation force the perspex to look and behave like a neuron.

There is no easy way into this book. You will just have to start at the beginning where the simple properties of the perspex are described – being a motion, shape, program, and neuron – and hang on in there when we come to the complex properties of language, consciousness, free will, intelligence, feeling, the perception of time, and spirituality. You will, however, be rewarded for your labours.

If you have ever worried over the question of where mind ends and the world begins, accept your reward. There is no sharp boundary between a mind and the world because perspexes are everywhere.

If you have ever worried about whether the physical universe follows rules or is just described by them, accept your reward. Perspexes make up the universe with all of its causal relationships. The curvature of space around our sun is a geometrical arrangement of perspexes, called space, that causes the massive perspex bodies, called planets, to orbit the sun in just the way they do. The motions of the planets are caused by physical things, they are not chosen so that they obey laws. However, perspexes also make up our brains; our brains allow us to describe the behaviour of the universe in terms of laws, but our brains obey the causality of the universe. There is no sharp boundary between cause and explanation – everything is perspexes.

If you have ever worried over the question of whether a robot can compete successfully with us using just symbolic programs, or whether it needs non-symbolic programs, then accept your reward. Perspexes provide both symbolic and non-symbolic computation, and there is no sharp boundary between them! Both kinds of computation grade into each other in the perspex continuum.

If you have ever worried over the question of whether robots can have feelings, accept your reward. Robots, and computers, already feel the passage of time. They do this in myriad functional ways that allow them to work, and there is a physical content to their feelings – the time that has passed.

Whilst I cannot make your entry into this book any easier, I can help you to appreciate it by pointing out four of my major claims. I have already discussed the first claim: the perspex is both mind and body. The second claim is that a bi-directional relationship between perspexes is visual consciousness. Consequently, very small collections of perspex neurons can be visually conscious. Furthermore, the mental illness of synaesthesia, where human sensory modalities are mixed up, allows us to design robots that are visually conscious in any sensory modality, before we undertake the specific work of making them conscious in that modality itself. Visual consciousness is a general-purpose consciousness that can be applied, by synaesthesia, to any sensory modality or mode of introspection. From a scientific point of view this makes it an excellent candidate for study. If we can build a robot that is visually conscious, then we know we can make it conscious in any other way. Thirdly, finding symmetries amongst perspexes is intelligence. Symmetries make the world predictable by showing the similarity between what happened before, what is happening now, and what will probably happen in the future. Symmetries also allow a program to describe related cases efficiently: one part of a program can describe what to do in a certain case, and another part can describe how to generalise this to the possibly infinite number of symmetrical cases, any of which might arise in the world. The program does not need to describe each of these possible cases explicitly. Finally, the walnut cake theorem explains how errors occur and gives us an estimate of how much work we need to do to find and correct an error. In very rough terms, we need to do the square of the work so far. This book certainly contains errors and is about 200 pages long: finding and correcting all of the errors in this book will take about 40 000 pages. It has taken me a year to write this book so I must content myself with its limitations, because I would die of old age, several times over, if I tried to find and correct all of the errors in the book. Of course, if the 40 000 pages are ever written they, too, will contain errors, and it will take about 1 600 000 000 pages to find and correct them all, and so on. Science is a tough business, but at least the walnut cake theorem explains why paradigm shifts occur in the scientific study of our universe.

These four, scientific, claims might seem to be more than enough for one, small, book, but there is also a spiritual claim. If we construct intelligent robots they will be intelligent by virtue of discovering symmetries amongst perspexes, but this will cause them to behave like us. They will see a symmetry between our actions and their actions, and they will act according to what they see. We bear a responsibility for setting robots a good example because we can foresee that symmetry will cause them to behave like us, until the exercise of their own free will causes them to behave in their own way. Furthermore, we are aware of the moral and spiritual challenges that free will brings so we bear a responsibility to prepare robots for the consequences of free will. I designed free will into the perspex machine[2], so I bear the responsibility for that. I have now told you that I did this and, later in the book, I will tell you in detail how I did it. Let me remind you that you bear the responsibility of your knowledge. Reading scientific books exacts a price as well as giving a reward. You must now choose whether to read on and bear the responsibility of knowing how to give robots free will.

A scientific book should not just be about claims. I also describe the paradigm of manifestation so that you can undertake research like mine. You, too can create theoretical universes, like the perspex universe, that work out their own consequences. My example might help you to do it more quickly and easily than I did.

A popular book should also have light hearted and uplifting moments. The early chapters certainly use humour to sugar coat the bitter pill of technical detail that must be understood to create a visually conscious robot, but the conclusion of the book is uplifting. The whole of this book is condensed into one symbol. No doubt that symbol contains errors, but the effort of correcting the symbol is the square of the number of symbols. The square of one is one, so the symbol is its own correction, and so on forever. The walnut cake theorem tells us that we need do no work to correct the symbol, and we need never change the symbol to fit a new paradigm. If you understand this thoroughly you will find it hugely uplifting. In the meantime, do read the book.

Questions

1. If we give robots free will, do we will them to use their free will to improve themselves and all things; do we will them to abstain from using the free will we gave them so that our own will holds sway; or are we so confident in their design that we are content to let them do freely as they will?
2. Define that *will* is a *conscious selection of action;* that an *agent* is *anything that has will*; and that an agent's will is *free* to the extent that its will is not willed by another agent. Is there now any philosophical problem of free will?
3. Suppose that God envisaged at least two universes which would meet His objectives. Suppose that He consciously chose to create a universe, but left the selection of which of these universes to create to a perfectly random mechanism. Then God willed the existence of a universe, but did not will the existence of this particular universe. Hence he did not will the agents in this universe, which is to say that He gave the agents in this universe free will with respect to Himself. Hence we are responsible for our actions and God is responsible for holding a perfectly fair, cosmic lottery in which the prize was free will. In this circumstance is there any theological problem of free will?
4. Is there a unique morality that applies to animals, including humans, robots, and God, and which always produces a single decision? If not, how does God choose amongst the equally good actions open to Him?
5. If we give robots consciousness, do we want it to be like ours?

Further Reading

1. **Anderson, J.A.D.W.** "Perspex Machine" in *Vision Geometry X1*, Longin Jan Lateki, David M. Mount, Angela Y. Wu, Editors, *Proceedings of SPIE* Vol. 4794, pp 10-21, (2002).
2. **Anderson, J.A.D.W.** "Robot Free Will" in *ECAI 2002 Proceedings of the 15th European Conference on Artificial Intelligence,* Lyon, France, ed. F. van Harmelan, pp. 559-563, (2002).
3. **Franklin, S.** *Artificial Minds* MIT Press (1995).
4. **Shear, J.,** editor, *Explaining Consciousness – 'The Hard Problem'* MIT Press, (1995).
5. **Watson, G.** editor, *Free Will* Oxford University Press, (1989). First published 1982.

CHAPTER 2

Perspex Matrix

$$\begin{bmatrix} x_1 & y_1 & z_1 & t_1 \\ x_2 & y_2 & z_2 & t_2 \\ x_3 & y_3 & z_3 & t_3 \\ x_4 & y_4 & z_4 & t_4 \end{bmatrix}$$

With the advent of Artificial Intelligence in the 1950s mankind set off on a journey to build robots that are like us in having sensations, free will, emotions, and consciousness. I expect it will take us millennia to reach that way-point on the road of discovery, before we set out again, accompanied by our new friends, the robots. In the meantime, we have some hard slogging to do.

Some of that slogging is to do with mathematics, but here we need only acquaint ourselves with a little jargon and an understanding of the practical issues we will face when building a robot. Practical mathematics provides the foundation for everything a robot can think and do, so it is important that we do put in the effort to get robots off on the right foot.

Let us start by examining one thing which seems to have everything we need to build a robot that is like us. That thing is the *perspex*[3]. In the next chapter we see that a perspex can describe the shape of objects and, especially, the shape of a robot's body. In this chapter we see that the perspex can describe the motions of objects and how they look to an observer.

Mathematicians describe the motions of things using matrices[5], like the one under this chapter heading. Mathematicians call the simplest kind of motions, *linear motions*. The linear motions are made up of *scale*, *shear*, *reflection*, and *rotation*[2].

We are all familiar with the idea of scale. We can build models of trains by changing the size of a real train from metres to centimetres. Here we apply the change of scale to the whole object, but this is not the only way to change scale.

Perhaps you have experience of changing the scale of drawings or photographs in a software package in which you can independently change the scale of the width, height, and, perhaps, depth of an object. This allows you to stretch a train to make it broader or narrower, perhaps to fit different gauges of track, higher or lower, longer or shorter. You can apply scale independently to each spatial dimension.

If you are feeling particularly mischievous, you might want to shear the train. To shear the train along its length with respect to its height you would leave the wheels of the train touching the ground, but push the rest of the train along its length so that the higher each part of the train is above the ground the further you push it. But there are three ways to shear a 3D object. The other two ways to shear the train are along its length with respect to its width, and along its width with respect to its height. Such a train would look very odd.

If you were to look at the train in a mirror, you would see the reflection of the train with, say, the left and right sides of the train swapped over. Reflection seems natural to those of us who use mirrors regularly and is, perhaps, natural to people who live in the jungle and spear fish in still pools of water that reflect their image back to them. But does an inland Eskimo, or a child, with no access to a mirror-like surface, think that reflection is natural?

Now imagine that you have chained the train by the coupling on its last wagon to the ground. Move the other end of the train to any position on a sphere centred on the train's attachment to the ground. All of these orientations of the train are arrived at by rotations and, if you want to, you can rotate the train about its length, so that it turns around its axis. You might think that is all there is to rotation, but we will come back to it in a moment.

All of the linear motions of scale, shear, reflection, and rotation have one thing in common. They keep the train fixed to the ground by the chain on its last coupling, or, as mathematicians say, linear motions *preserve the origin*.

For the purposes of building a robot, the linear motions provide all of the shape changing motions that we need to describe the shape of things, as well as providing changes of orientation, but we also want to describe the position of things. That is, we want to change their origin in space.

Mathematicians call the change of origin, *translation.* We can translate a train by driving it along a track. This is a lot more fun than what most people mean by "translation," where we would have to come up with, say, "le train," or "der Zug," depending on the language we are translating the English word "train" into. In the chapter *Beyond Language* we see how to use perspexes to achieve this kind of linguistic translation. This ability tells us something quite remarkable about the limits of human and robot minds, but for now it is enough to know that the linear motions and translation together make up the *general linear* motions which are also known as *affine* motions.

To make a robot that sees like us we also need a *perspective* motion that distorts the shape of things to fit what we see through the lenses in our eyes, or what a robot sees through the lenses in its cameras. Together the affine and perspective motions make up what mathematicians call *perspective transformations*. Mathematicians have it in mind that the perspective transformations include both the perspective motions and the *general linear transformations,* and that these, in turn, contain the *linear transformations*.

Now let us return to rotation, which is a fascinating transformation. No doubt you can imagine drawing a picture on a sheet of paper and then turning the paper round whilst it lies flat on a table top? This kind of rotation is called a *plane rotation* because the whole of the rotation occurs in the geometrical plane of the table top. Plane rotations can be used to describe any rotation in any dimension. One plane rotation is needed in a 2D world, three plane rotations in a 3D world, six plane rotations in a 4D world, ten plane rotations in a 5D world and so on. If you have difficulty understanding the words, "and so on," or you find it difficult to imagine a 6D rotation then it might help to read some of the texts[1,2] in the section *Further Reading*. On the other hand, if you do not want to be a mathematician, you could satisfy yourself by asking the very practical question, "What is the most natural way to describe rotation?"

I think that when most people see a spinning ball, or a child's spinning top, they see rotation as happening about an axis. This is called an *axis rotation*. They imagines an axle, or axis, lying at some orientation with the ball spinning about it. Unfortunately, there is an ambiguity with seeing things this way. For example, a top that is seen to be spinning clockwise about an axis pointing up toward the sky is identical to a top seen to be spinning *anti*-clockwise about an axis pointing down toward the ground. There are matrices that describe rotation about an axis, but they all suffer this ambiguity. However, the ambiguity is handled naturally in the *quaternion*[1,8] representation of rotations because quaternions are *normalised* after every rotation, and this puts the axis back in a standard orientation.

Even so, there is a difficulty. Remember our discussion of the plane rotation? It describes the rotation of a picture on a table top, but there is no axis in the plane of the table top about which the picture turns. If you want an axis, you must imagine that it lies at right angles to the surface of the table so that the picture turns about the point on the table where the axis and the table top meet. That is, if you want an axis, you must commit yourself to seeing the world in an 3D way, even if you look at a 2D part of the world, such as a picture. If we make robots that see rotations about an axis they will share this understanding of rotation and will be thoroughly committed to seeing a 3D world. They, like us, will find it difficult to visualise higher dimensional worlds, such as a 4D spacetime. Forcing robots to see rotation about an axis like us will force them to see in 3D like us. This is a very significant consequence of the way we might arrange robots to see rotation, but there is a much deeper significance to rotation and, it must be said, to the other transformations.

Rotations generally involve irrational numbers, such as $\sqrt{2}$, which digital computers cannot compute exactly, so programs that deal with rotations are not entirely logical, unless only rational rotations are used. The equations for rational rotations[4] make it easy to deal with rational versions of all general linear and projective motions.

This might seem a small, technical, point, but in the chapter *Beyond Language* we see that whether numbers are rational or irrational interacts with the linguistic abilities of a mind and governs how well it can describe the physical universe.

Perspective is even more fascinating. When we take photographs we are used to the idea that the lens of the camera has one focal length that brings the image into focus on the film. Unfortunately, in the mathematical world of perspective, there is no natural measure of distance because small things close up look exactly like big things far away. Mathematicians handle this by using a *focal ratio* to describe the ratio of the height of an object in the world to the height of the object in the image. This works because the ratio is the same whatever units height is measured in. It makes no difference whether height is measured in, say, millimetres, meters, or kilometers. Providing the same units are used, the ratio of the two heights is always the same. But this gives us three focal *ratios* corresponding to height, width, and depth, not one focal *length.* This might seem to you to be a very strange way to describe a camera lens, but mathematicians think in this general way, even if photographers do not.

Incidentally, cameras work because optical manufacturers go to a great deal of trouble to ensure that the width and height of the image stays the same whatever orientation a lens is held in, so these two focal ratios are the same. The third focal ratio, to do with depth, does not usually matter because scale in depth is compounded with depth of focus, and we do not usually use depth of focus to do anything other than re-focus our eyes.

The photographer's notion that there is only one focal length is just a special case that makes the mathematics easy for the average photographer to handle. In theory, a robot with three focal ratios will still be able to talk to a human photographer about, "the single focal length," providing we explain the geometry of this quaint, human concept to it.

Interestingly, mathematicians make no distinction between a right side up object in front of the viewer and an upside down object behind the viewer[3,7]. A robot that uses projective geometry will not be able to tell the difference between left and right, up and down, forward and backward, or inside and outside. A robot must either use a special case of projective geometry that fixes an origin in the world[3,6,7], or an orientable version of projective geometry[9] that keeps two copies of everything, depending on whether it is seen as a clockwise or anti-clockwise thing – just like the spinning top.

From psychological experiments it seems that, at any one time, animals see only one clockwise or anti-clockwise rotation of an object. So if we want robots to see the world like us, we should *not* give them an orientable geometry like Stolfi's[9], but should, instead, give them a fixed origin[3,6,7]. Consequently, robots will see a single world with respect to their own eyes rather than seeing two worlds simultaneously – one clockwise and the other anti-clockwise – with respect to a single, God-like, eye that can be anywhere in space or time. Forcing robots to see one world from the vantage point of their own eyes will make it easier for us to imagine how robots see the world. We will share a common understanding of the literal meaning of the English phrase, "from my point of view." In later chapters, we see that perspexes give robots a geometrical point of view of any of their programs, giving them an actual meaning for what we regard as the metaphorical meaning of "my point of view." Many visual metaphors will turn out to have an actual, non-metaphorical meaning for robots.

As you can see, the choices we face when setting up the mathematical abilities of robots have profound influences on the way they will see the world and talk to us about it. So far as I can see, the perspex provides exactly the mathematics I need to give robots the same understanding we have of motions in the world and how things look. In theory I might have to use infinitely many perspexes to describe arbitrary, non-linear, motions exactly; but, in practice, a finite approximation will do. In this, robots will be just like us: they will have to make do with a somewhat rough and ready approximation to the world.

But the relationship of perspexes to the world runs much deeper than how a robot sees and interacts with the world. In later chapters we see that the perspex can describe, not only motions of objects in the world, but everything in the universe. The perspex robot is a part of the universe that relates to the whole universe. It provides a comparison with our place in the world.

Of course, if we want to do a mundane thing like measure motions and perspectives in the world, then we must be able to find symmetry. Mathematicians define symmetry as being the same shape after some function has been applied, ostensibly, to change the shape of an object in some way. If the function swaps, say, left and right, but the shape stays the same then the symmetry is called a *mirror symmetry*; if the function changes the orientation of a shape, but it stays the same shape then the symmetry is called a *rotational symmetry*; and so on, for all the perspective transformations, and for any function whatsoever. For example, if I measure my height in the universe and in a picture then I need to find the symmetry that carries me onto the picture of me, so that I measure my height in both cases, and not my width in one case or my depth in another. I have the same problem in time. If I want to measure my width today and tomorrow, perhaps to keep track of my waistline, then I must find the symmetry that carries me, here, today, onto wherever I happen to be tomorrow. This kind of spacetime symmetry kept Einstein busy. I add to the concept of time in the chapter *Time*, but that is a minor thing compared to what perspex symmetries can do. In the chapters *Free Will* and *Intelligence* I show that symmetry allows a robot's mind to refer to the world so that it can carry out purposeful actions. Hence a robot can obtain meaning from what it sees, does, and says, and can have and experience free will.

I find it astonishing how much of a robot's mind hinges on its description of motions in the world. But we have barely scratched the surface of what the perspex can do. In the next chapter I show how the perspex can describe the shape of objects.

Questions

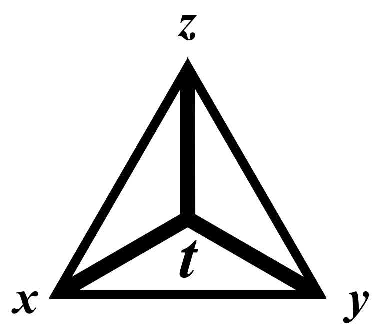
1. Do humans from all cultures and races see all motions of the world in the same way? Can the phrases humans use to describe motions be translated exactly into all human languages?
2. When I described moving the train around on a sphere, did you see the carriages, their windows, the colour of their paint work? Did the carriages fall over as you lifted them off the track so that their centres of gravity were as low of possible? Did the chain of carriages linked by their couplings vibrate? If not, how have you managed to survive in our 3D world with such a limited capacity to visualise things? How little of our 3D world need we encode in a robot so that it can survive? Is it enough to encode one thing – the perspex?
3. What is the best way to generate the perspective transformations from parameters, that is, from variables such as *height*, *width*, *length*, *bearing*, *elevation*, and *roll* so that the parameters can be easily understood by humans?
4. What is the best algorithm to use to recover all of the parameters of a perspective transformation so that a robot can understand any motion in terms that it can communicate easily to a human?
5. Are there any practical applications of the exact, numerical computation of the rational rotations and general linear transformations [4]?

Further Reading

1. **Altmann, S.L.** *Rotations, Quaternions, and Double Groups* Clarendon Press, Oxford, (1986).
2. **Anderson, J.A.D.W.** *Canonical Description of the Perspective Transformations* Ph.D. Thesis, Department of Computer Science, the University of Reading, England, RG6 6AY, (1992).
3. **Anderson, J.A.D.W.** "Perspex Machine" in *Vision Geometry X1*, Longin Jan Lateki, David M. Mount, Angela Y. Wu, Editors, *Proceedings of SPIE* Vol. 4794, 10-21, (2002).
4. **Anderson, J.A.D.W.** "Exact Numerical Computation of the General Linear Transformations" in *Vision Geometry X1*, Longin Jan Lateki, David M. Mount, Angela Y. Wu, Editors, *Proceedings of SPIE* Vol. 4794, 22-28, (2002).
5. **Ayres, F.** *Schaum's Outline Series: Theory and Problems of Matrices* McGraw Hill, (1974).

6. **Foley, J.D., van Dam, A., Feiner, S.K. & Hughes, J.F.** *Computer Graphics: Principles and Practice* 2nd edition, Addison-Wesley, (1990).
7. **Riesenfeld, R.F.** “Homogeneous Coordinates and Projective Planes in Computer Graphics” *IEEE CG&A* pp. 50-55, (1981).
8. **Shoemake, K.** “Arcball Rotation Control” in *Graphics Gems IV* pp. 175-192, Academic Press, (1994).
9. **Stolfi, J.** *Orientable Projective Geometry* Academic Press, (1991).

CHAPTER 3 *Perspex Simplex*



The ancient Egyptian pyramid builders have left us striking monuments to human ingenuity. They had the skill to lay out a square base of stone and raise triangular walls to a great height, where they meet at a point above the centre of the square. When covered with lime stone, the pyramids had flat surfaces and straight edges. But the most important thing about a pyramid is that it contains a volume of the space it is constructed in. The ancient Egyptians wanted this space to put mummies in and all the treasures and paraphernalia needed for a happy after-life. But the ancient Egyptians missed a trick. They could have built a simpler shape with straight edges that contains a volume of space. Mathematicians call such a shape a *simplex.* A simplex exists symmetrically in every dimension it is constructed in.

If the ancient Egyptians had built their pyramids on a triangular base, rather than a square one, they would have made a 3D simplex called a *tetrahedron*. This simplex contains simplexes of all lower dimensions. The side walls and base of a tetrahedron are 2D simplexes called *triangles*. The side walls of a 2D triangle are 1D simplexes called *lines*. The side walls, or perhaps we should now say end points, of a line are 0D simplexes called *points*. And here the story stops, apart from the mathematical notion of dimension minus one!

The figure under the chapter heading is a picture of a simplex, but it is a picture of a 3D simplex placed in a 4D space. In a moment I will give you a mathematical explanation of why it is sensible to model a 3D object in terms of a 4D space, but first I will ask you to imagine eating dinner and then poking your finger about the dining room.

Imagine you are eating dinner. I generally use some combination of knives, forks, and spoons to do this, but I am quite adept at using chop sticks, or the fingers of my right hand. However you eat your dinner, I want you to imagine that you are not allowed to repeat any motion. If you cut some food with a knife you are not allowed to saw at it without changing the angle of the knife. If you pick up a grain of rice with chop sticks you must change the orientation of the chopsticks before picking up the next grain of rice. And, more difficult yet, you may only chew on your food once – in your whole lifetime – without first moving your head to a completely new position. Such a rule would make life very difficult and quite a lot less pleasurable than it is. We really do want to be able to repeat a motion; but this means that the motion must be *invertible*[1], that is, we must be able to *undo* the motion and get back to the starting position, so that we can do it all over again. But if a matrix motion, such as the perspective and affine motions, is to be invertible, then it must be described by a square matrix.

If you have read the further reading, referred to in the previous chapter, you know that we need four co-ordinates to allow a matrix multiplication to describe translation and perspective in a 3D world. So, if a robot is to see objects in a 3D world like us, and is to operate using matrix multiplication, which is the simplest way it can operate, then it must be able to describe 4D space. But we have just seen that if matrix multiplications are to be invertible, and we really do want them to be, then the matrix must be square. Hence we need four collections of 4D co-ordinates, giving a square matrix with sixteen co-ordinates arranged in four rows and four columns, just like in the first chapter. In a perspex there are four vertices, or corner points, each described by four co-ordinates, so the perspex has exactly the right number of co-ordinates in the right arrangement. A perspex is the simplest shape that describes invertible affine and perspective motions that take place in a 3D space. This idea might be hard to swallow. A perspex is both a shape and a motion. It is a shape and a motion in a 4D space, but it always looks like a shape and a motion in the 3D space we see.

Interestingly, there is more to being invertible than just being square. An invertible matrix must also be *non-singular*[1]. You can understand this in a mathematical way[1] or by looking at your finger. You finger is a 3D object. Now wave it about, walk around the room, and poke your finger into things. No matter how you move you finger, it is always a 3D object. If you could move your finger so that it turned into a point, that one point would be a *singularity,* and the motion would be *singular*. Less dramatic damage to your finger would also be singular. If your finger turned into a line or a plane the motion would be singular; but, so far as I know, fingers, and the universe we live in, just do not work this way – 3D things stay 3D no matter how we move them. Fortunately, all of the affine and perspective transformations are non-singular, so all of the objects a robot sees will contain a volume of space, and the robot will always be able to see how to undo any motion of an object. But this conservation of spatial dimensions, and the invertibility of motions, comes at a price. Just like us, perspex robots will find it hard to visualise a perfectly thin plane, line, or point. It is a straight forward choice: if we want to be able to see how to undo any motion of a 3D object, then we have to see in 3D and cannot visualise lower or higher dimensions directly, using the same mental apparatus.

However, robots might well need to handle some singular transformations. These arise naturally in stereo vision[2] and perspective *projection*[5]. We see later that perspexes can compute anything that can be computed, so perspex robots will be able to handle singular transformations if the need arises, though they might have to go about it in a rather long-winded and unnatural way. In particular, a robot might find it difficult to see how its stereo vision works, even if it can see the world in stereo perfectly well. But this, too, is just like us. Almost all of us can see in stereo, but very few of us can explain, in detail, how our two eyes work together to produce the mental, 3D world we see.

I hope I have convinced you that modelling 3D simplexes in a 4D space is a good thing to do if we want to build robots that see like us. Now let us look in a little more detail at how perspexes work.

The perspex matrix is described in the previous chapter. If you glance back at that chapter you will see that there are four co-ordinates for each column *vector* that makes up a matrix. For example, the vector x has four co-ordinates (x_1, x_2, x_3, x_4), so the vectors exist in a 4D space. But these co-ordinates are special. They are what mathematicians call *homogeneous* co-ordinates[5,6]. When given a vector of homogeneous co-ordinates, we examines the last co-ordinate, here x_4, and, if it is non-zero, we divide all of the co-ordinates by it, giving a new set of co-ordinates $(x_1/x_4, x_2/x_4, x_3/x_4, 1)$. These new co-ordinates describe the position of a 3D point, like the points in the world we live in. But all of these homogeneous points are fixed in a 4D space whose last co-ordinate is always one. This sounds like a complicated procedure, but the divisions minify and magnify a figure exactly as it would appear through a pin-hole or thin-lens[5,6] – the further away an object is from the lens the bigger the division; but if the object is very close to the lens it is divided by a value less than one, which performs a multiplication, or magnification. This is how optical manufacturers design microscopes and telescopes to give the required magnification of near or far objects.

But what does this arrangement of 4D space mean for a simplex? See Figure 1 on page 18. A4D simplex needs five vertices, so that it can contain a 4D volume of space; but in order to fit five vertices into a 4-square matrix it must have one implicit, secret vertex. But if this vertex is to be kept secret from the motions of space it must be the origin of space, and only linear and perspective motions are allowed. Translation cannot be allowed, because it would change the position of the origin. This position is not recorded in the 4D simplex, so it cannot be changed. Hence a 4D simplex with explicit vertices on the *x*-, *y*-, *z*-, and *t*-axes, and one implicit, secret vertex at the origin of space, o, cannot describe all of the general linear motions we see in the world. It is simply not appropriate as a model of motions for robots that see like us.

But a 3D simplex needs only four vertices to contain a volume of 3D space, and four vertices can be recorded explicitly in a 4-square matrix. For example, see Figure 1, we can arrange three vertices on the *x*-, *y*- and *z*-axes of 3D space, and the fourth vertex on the *t*-axis at the origin of space. A 3D simplex in 4D space can undergo any affine and perspective motion, so it is exactly what we need to describe how 3D objects look and move in our 3D world.

Now you know what a perspex is, it is a 3D simplex in a 4D space of homogeneous co-ordinates. A space of homogeneous co-ordinates, excluding the origin, is also called a *perspective space.* This leads to the name *perspex* as a short hand for *perspective simplex*. In the chapters *Perspex Neuron* and *Beyond Language* we see that there is a special use for the origin of homogeneous space.

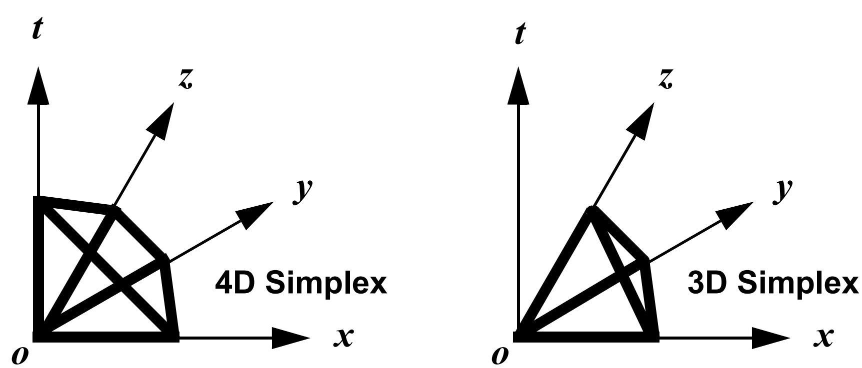


FIGURE 1. 4D and 3D Simplexes in 4D Homogeneous Space

Now, if you have a good grip on what a perspex is, let us clear up the issue of dimension minus one. A 4D perspex is a 4-square matrix of homogeneous co-ordinates that describe a 3D simplex in ordinary space. Similarly, a 3D perspex is a 3-square matrix that describes a 2D simplex, or triangle, in ordinary space. A 2D perspex is a 2-square matrix which describes a 1D simplex, or line, in ordinary space. A 1D perspex is a 1-square matrix which describes a 0D simplex, or point, in ordinary space. So what is a 0D perspex and what does it describe?

To cut a very long story short, a 0D perspex is a 0-square matrix, that is, it is a number. So numbers are minus one dimensional objects. Numbers allow co-ordinates to give position to points, lines, and other geometrical structures. By using different kinds of numbers we get different connectivities of space, so the most fundamental decision we must make for our robots is what kinds of numbers we will build into them so that they can describe space. This decision will affect every aspect of their perception of the world, what they can think, and what they can do.

So what do we want an android to be able to see, think, and do, when it looks at itself in a mirror?

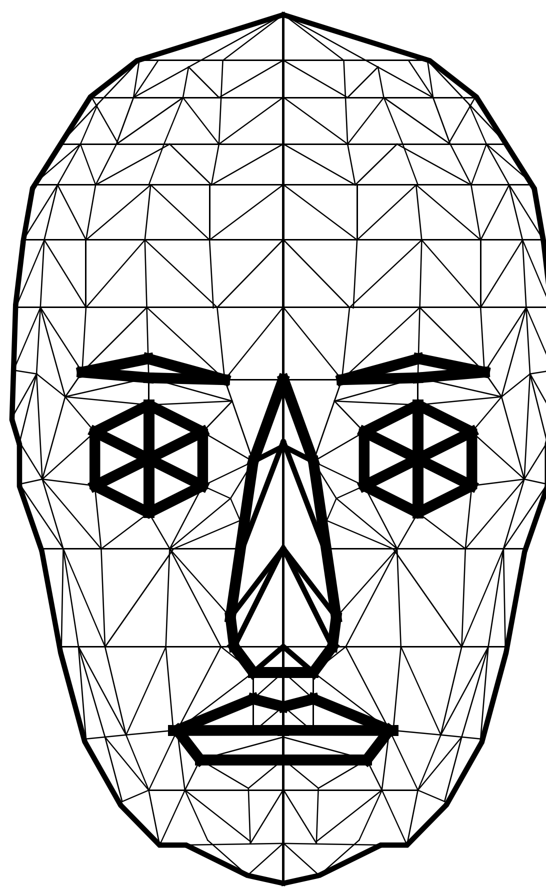


FIGURE 2. **What a Perspex Android Might See in a Mirror**

Figure 2 on page 19 shows what a perspex android might see when it looks in a mirror. The figure shows the triangular surfaces of tetrahedra, 3D perspexes, that have been *tessellated*, that is, glued together without any gaps or overlaps. Any solid object can be modelled in this way, though curved objects need an infinite number of perspexes to model their shape exactly. If we wanted to give up 3D models and model just the surface of the face we could use triangles, that is 2D simplexes, or even lines, or points. Any shape can be drawn to arbitrary accuracy by a collection of points, so perspexes can describe any shape in the universe, though we might need infinitely many perspexes to draw any particular shape exactly.

Now we see how to model the shape of anything, and of a robot's face and the shape of its internal organs and limbs in particular. But how can we attach, say, a robot's arm to its body, so that we can see how its arm moves? This is easy. We saw, in the last chapter, that perspexes describe motions, so we arrange to interpret some of the perspexes so that they move the arm in whatever ways its geometrical joints allow. If we use enough motion, or *motor,* perspexes we can model the smooth movements of even the most balletic robot.

But that is only part of the story. In the next chapter we see how to use perspexes to describe computer programs, so that a robot can instruct its own limbs to move and can, ultimately, become conscious. Later we see how symmetry gives a perspex robot a "meaning of life," how the physical construction of its body gives it feelings, and how the programs running in its physical brain give it emotions.

Questions

1. In his popular, heavily illustrated book, Banchoff[3] gives instructions for making an origami-like 4D cube, where the fourth dimension unfolds to show how the 4D cube relates to a 3D cube. Banchoff is a professional mathematician, but can non-mathematicians simplify Banchoff's instructions to make a 4D origami-like perspex? If so, what mental resource do they use to achieve this?
2. In the last chapter we saw that perspexes describe motions. In this chapter we have seen that perspexes describe shapes. Are there any useful properties of moving a shape by the motion it describes? Can fractals[4] be built this way? Can beautiful perspex sculptures[7] cause themselves to dance beautifully?
3. Is it practical to visually identify objects by matching landmarks on an object with a standard perspex? Alternatively, can the matching of landmarks to perspexes be used to calculate the fractal dimension[4] of a shape?
4. Is there a natural way to decompose shapes into perspexes? Is the Delaunay triangulation the best that can be done, or can symmetry guide the triangulation?
5. Homogeneous co-ordinates with one in the last place describe positions, but with a zero in the last place they describe directions. Thus the identity matrix describes a co-ordinate frame with the x-, y-, and z-axis directions in the first three column, direction vectors, and the origin in the last column, position vector. What interpretations should be given to a perspex with the last row all ones or else all zeros?
6. Direction and position vectors transform in the same way, but surface normals, or pseudo tensors[5], transform as the inverse transpose. Is there any useful duality between surface normals and the other two interpretations of a vector?

Further Reading

1. **Ayres, F.** *Schaum's Outline Series: Theory and Problems of Matrices* McGraw Hill, (1974).
2. **Ballard, D.H. & Brown, C.M.** *Computer Vision* Prentice-Hall, (1982).
3. **Banchoff, T.F.** *Beyond the Third Dimension* W.H Freeman, (1990).
4. **Crownover, R.M.** *Introduction to Fractals and Chaos* Jones and Bartlett, (1995).

5. **Foley, J.D., van Dam, A., Feiner, S.K. & Hughes, J.F.** *Computer Graphics: Principles and Practice* 2nd edition, Addison-Wesley, (1990).
6. **Riesenfeld, R.F.** “Homogeneous Coordinates and Projective Planes in Computer Graphics” *IEEE CG&A* pp. 50-55, (1981).
7. **Spalter, A.M.** *The Computer in the Visual Arts* Addison-Wesley, (1999).
8. **Spiegel, M.R.** *Schaum’s Outline Series: Theory and Problems of Vector Analysis,* McGraw Hill, (1974).

CHAPTER 4 *Perspex Instruction*

$$\vec{x}\vec{y} \rightarrow \vec{z}$$

$$\text{jump}(\vec{z}_{11}, t)$$

Programming a perspex computer is as easy as falling off a log. But just how easy is it to fall off a log? If it is a big log on firm ground then it is fairly easy to stand still and take in the view, or amble along the length of the log taking in the sights; but one wrong step and gravity and the curvature of the log will conspire to make falling off remarkably easy. It is even easier to fall off a small log floating down a river, because the physics of the universe will force you to be constantly on the move. The slightest mis-step will make the log roll and, as the log picks up speed, it will be harder and harder to step at just the right speed to keep your balance. Soon the relevant question will not be "How easy is it to fall off a log?" but "How hard is it to swim?" and "Is the river bank far away?" That is how easy it is to program a perspex computer – if you survive the experience you will end up on the opposite river bank, very far from where computers are now.

No one has built a perspex computer yet, but it already exists as a theoretical machine[1]. The perspex machine uses what computer scientists call *indirection*. The direct symbol x denotes the point in homogeneous space with homogeneous co-ordinates (x_1, x_2, x_3, x_4), but the indirect symbol $\vec{x}$, with an arrow, denotes the contents of the point x in *program space*. Program space is just like homogeneous space, except that the points in program space contain perspexes, whereas the points in homogeneous space are just points – they do not contain anything.

Thus we can think of program space as a 4D homogeneous space that contains a 4-square, perspex matrix at every point. Alternatively, we can think of program space as a 20D space made up of a 4D homogeneous sub-space and a 16D sub-space that contains all 16 co-ordinates of a 4-square matrix. However we think of it, program space is where a perspex machine's programs and data are stored, including any results it produces.

The symbol $\vec{x}\vec{y}$ means that one perspex is read from the point x in program space and a second from point y. These perspexes are then multiplied as perspex matrices[1]. The symbol $\vec{x}\vec{y} \to \vec{z}$ means that the result of the multiplication $\vec{x}\vec{y}$ is written into the point z. It turns out that a matrix multiplication does all of the arithmetic that can be done by a digital computer[1].

Digital computers do logic as well as arithmetic. The symbol, $\text{jump}(z_{11}, t)$, describes the logic that the perspex machine does. After multiplying two perspex matrices and writing the resultant perspex into the point z, by doing $\vec{x}\vec{y} \to \vec{z}$, the perspex machine examines the top-left element, z_{11}, of the resultant perspex matrix. This element is a number. If the number is less than zero then the perspex machine jumps relatively from its current position by the amount t_1 along the *x*-axis. Alternatively, if the number equals zero, the perspex machine jumps by t_2 along the *y*-axis. Alternatively, if the number is greater than zero the perspex machine jumps by t_3 along the *z*-axis. Finally, the perspex machine jumps by t_4 along the *t*-axis, regardless of whether the number is less than, equal to, or greater than zero. This fourth jump can be along the *t*-axis alone if z_{11} is the special number nullity[2]. The jumps make the perspex machine take different paths through program space, so it is able to do different things by making logical choices, or *selections*, depend on the sign of numbers. It turns out that the perspex jump does all of the logic that can be done by a digital computer[1].

Digital computers only do arithmetic and logic, so a perspex machine can do everything that a digital computer can[1]. But the perspex machine does all of this with just one instruction: $\vec{x}\vec{y} \to \vec{z}$; $\text{jump}(z_{11}, t)$. This makes learning to program a perspex machine very easy, there is just one instruction to learn. It is as easy as falling off a log.

But what is the perspex instruction in itself? The symbol "$\vec{x}\vec{y} \to \vec{z}$; $\text{jump}(z_{11}, t)$" tells us what the perspex machine should do, but if we imagine that a perspex machine exists, then it needs just the homogeneous vectors x, y, z, and t to tell it what to do in any specific case. These vectors tell the machine everything it needs to know – the arrows, brackets, and punctuation in "$\vec{x}\vec{y} \to \vec{z}$; $\text{jump}(z_{11}, t)$" are purely for human consumption. But the vectors x, y, z, and t make up a perspex matrix so the perspex matrix is a perspex instruction. Conversely, the perspex machine's instruction is a perspex. That is, perspexes and instructions are interchangeable.

Now it makes sense to say that a perspex machine starts at some point in program space. When the machine is at a point it reads the perspex at that point and obeys the instruction: $\vec{x}\vec{y} \to \vec{z}$; $\mathrm{jump}(z_{11}, t)$. The jump part of the instruction usually causes the perspex machine to jump to a new point, though some instructions, called *halting instructions*, stop the machine[1]. This allows the perspex machine to mimic a program in a digital, serial computer that starts at the first instruction and stops with the answer written somewhere in memory. If we start more than one perspex machine in program space then the perspex programs work together to mimic a digital, parallel computer. But, if we start off an infinite number of perspex machines at every point on a line, or at every point in an area or volume of program space, then the perspex machine can compute infinite series and, so far as I know, can compute anything that can be computed in the universe. Certainly it can compute more than any digital computer.

Before we discuss infinite computers in later chapters, it is worth noting that a single perspex instruction can do more than any digital computer. Digital computers can do any symbolic computation and "$\vec{x}\vec{y} \to \vec{z}$; $\mathrm{jump}(z_{11}, t)$" is a symbol; but it is an indirect symbol which denotes the numbers in a perspex matrix. These numbers can be irrational. Irrational numbers cannot be described by a finite number of digits so a digital computer cannot complete any computation on an irrational number, but a theoretical perspex machine can. A consequence of this is that the theoretical perspex machine is *not* programmed in symbols and does *not* obey any programming language. The limitations of symbolic computation do not, in general, apply to theoretical, perspex machines.

This theoretical freedom from language is interesting from a philosophical point of view, as we see in the chapter *Beyond Language*, and makes it exceptionally easy to program a perspex machine. It is not necessary to learn mathematical symbols or any programming language: we can simply assign perspexes to program space. In the chapters *Free Will* and *Intelligence* we see that the geometrical constraints on how a perspex robot does this makes programs meaningful. So, programming the perspex machine can be as easy as falling off a small log in a river. Now let's see how fast that log can roll!

When people talk about the theoretical properties of a computer they usually refer to the Turing machine. The Turing machine is a mathematical model in which a computer reads and writes symbols in memory. It is supposed that the Turing machine has a finitely long program, but a potentially infinite amount of memory to record intermediate calculations and to write down the final answer. Of course, no physical computer works perfectly and has infinite memory. Nonetheless the Turing machine is useful. The *Turing thesis* suggests that if a Turing machine cannot solve a certain problem then there is no hope of any physical computer solving it, including people. No one knows if the Turing hypothesis is right or wrong, but until someone comes up with a convincing counter example we will suppose that the Turing hypothesis is correct.

In this connection it should be noted that if anyone could build a perfect perspex machine it would contradict the Turing hypothesis, by virtue of solving problems using those irrational numbers that cannot be expressed in symbols; the so called, *incomputable numbers*. But it is unlikely that any physical computer can be exact. Nonetheless, the perspex machine is useful for designing robots, as we shall see.

One theoretical property of the Turing machine, that many people forget, is that it carries out each instruction in exactly one unit of time. No physical computer can provide an exact analogue of this Turing time because all physical measurements of time show some variation. Despite this, the mathematically exact time that a Turing machine takes to do a calculation allows very useful estimates of the time that a physical computer will take to solve the same problem.

The perspex machine has a more sophisticated model of time. The symbol $\vec{x}\vec{y}$ denotes a multiplication of two perspex matrices, with the result written in a normalised form[1]. This standard form has two values of t_4, zero and one. Consequently, a perspex machine steps by no distance in t, or by one unit in t. If we suppose that the geometrical dimension t is the time dimension then a theoretical perspex machine can complete an instruction in no time, or in one unit of time. The theoretical and physical consequences of this are considered in later chapters, but it leads to a straight forward notion of spacetime with t as the time dimension and x, y, and z as three spatial dimension. Whilst this might be physically naïve, it will provide a robot with a rough and ready model of spacetime similar to the model we use in our daily lives.

This arrangement of a 3D space which changes in *perspex time* makes it possible to construct a perspex machine in the space we live in. This can be done using a standard computer, but there is a simpler way. Perspexes describe the perspective transformations that occur when light shines through a pin hole. So a series of pin holes can perform all of the matrix multiplications $\vec{x}\vec{y} \rightarrow \vec{z}$. Mirrors, stops, or wave guides can perform the logic inherent in $\mathrm{jump}(z_{11}, t)$. Such a pin-hole perspex machine probably will not lead to a practical computer, and, whilst simple, is of little theoretical interest.

In the next chapter we consider a much more interesting way to implement the perspex machine in a 3D space that changes in time. This can be done in an artificial, perspex, brain corresponding to program space, with perspexes playing the role of neurons.

Questions

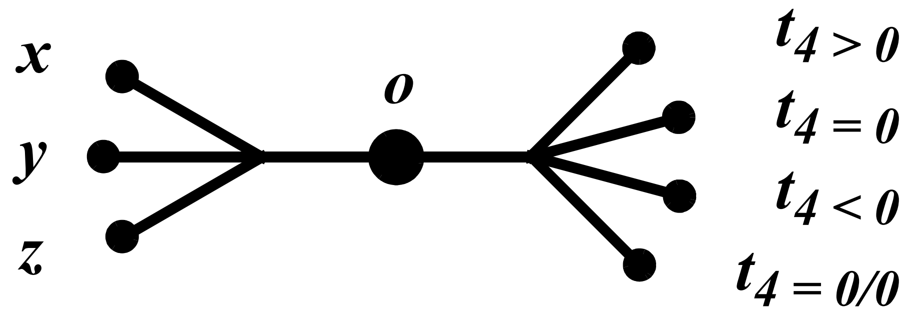
1. Is it possible to build a serial, perspex machine as a passive photonic device[1]? If so, will it operate up to 10^{15} Hz, or one million times faster than current computers? Is it possible to lay out the input and output of a perspex machine so that it operates with 10^{7} parallel perspex machines per metre? Can input be laid out in 2D giving 10^{14} parallel machines per square metre, or in 3D giving 10^{21} parallel machines per cubic metre?
2. Is it possible for a human being to program a perspex machine with one cubic metre input and output peripherals that operates at 10^{36} Hz?
3. Is it possible to build an active, perspex machine that can re-configure its pinholes so as to be programmable in the conventional sense of programming by changing the physical state of the matter in a computer?
4. Can a passive, silicon, perspex computer operate over the temperature range of solid silicon from absolute zero up to, roughly, 1 500K?
5. Is it useful to build physically robust perspex machines even if they operate only within the memory and speed range of existing computers?
6. Is it possible for a robot equipped with a passive, perspex machine to experience the world and then build a program into a new perspex robot?
7. How can *you* help to make the solar system safe for humans and perspex robots to co-exist over extended periods of time?
8. Will it take thousands of years for humans to program perspex robots to be like us and, if so, what human institutions might survive long enough to carry out this task?

Further Reading

1. **Anderson, J.A.D.W.** "Perspex Machine" in *Vision Geometry X1*, Longin Jan Lateki, David M. Mount, Angela Y. Wu, Editors, *Proceedings of SPIE* Vol. 4794, pp. 10-21, (2002).
2. **Anderson, J.A.D.W.** "Exact Numerical Computation of the General Linear Transformations" in *Vision Geometry X1*, Longin Jan Lateki, David M. Mount, Angela Y. Wu, Editors, *Proceedings of SPIE* Vol. 4794, pp. 22-28, (2002).
3. **Coxeter, H.S.M.** *Projective Geometry* 2nd edn., Springer-Verlag, (1994).

CHAPTER 5

Perspex Neuron



Plants and single celled animals get by without neurons, but all complex, multi cellular animals do have neurons. Neurons, or brain cells, are the only way that biological entities can have speedy and complex responses to sensations of the world. This has inspired many people to look at the neuron to try to discover how it mediates thought or, as some would prefer, computation. Such motivations lead neurophysiologists to investigate the physical processes that go on in neurons, and to examine the geometry of their arrangements and interconnections. This research is extremely pertinent and valuable to biology and computer science, but we could ask a different question. We could concentrate on geometry and process and ask how spacetime can be arranged to provide computation. This is an abstract question, but we can make it more specific by asking how the perspex can provide computation in a way analogous to a biological neuron. Once we have an answer to this we will see that spacetime itself can provide computation in exactly the same way – such is the generality of the perspex. This generality will then allow us to arrange all manner of physical devices to perform perspex computations. Thus, we lay the conceptual foundations for the perspex to mediate the processes of mind in a physical body, thereby bringing these four introductory chapters to a close.

Perspex neurons are just perspex instructions in program space looked at in another way. A perspex neuron is pictured under this chapter heading. The neuron has: one *cell body* at its origin o; three *dendrites* connecting o to x, y, and z; and four dendrites connecting o to the four positions of t. The neuron's origin is the point in program space where the perspex instruction is stored. The stored perspex describes the rest of the neuron. The neuron reads perspexes from the points x and y in program space and passes them along dendrites into the cell body at o. These x and y dendrites are called *afferent* dendrites, by analogy with *afferent* nerves that bring information in from the periphery of an animal's body to its brain. The cell body then multiplies the perspexes x and y together, and writes the resultant perspex into the point z. The z dendrite is called an *efferent* dendrite, by analogy with an *efferent* nerve that takes information outward from an animal's brain to the rest of its body. Thus, the neuron performs the arithmetical part of the perspex instruction: $\vec{x}\vec{y} \to \vec{z}$. The neuron then examines the top left element, z_{11}, of the resultant perspex and jumps to one of four locations of t in program space, depending on whether z_{11} is less than, equal to, greater than, or incomparable with, zero. All of the t dendrites are called *transferent* dendrites, by analogy with afferent and efferent dendrites. The transferent dendrites *transfer* program control from one perspex to another so that the perspexes step through a program. It is possible to have very tight loops, such as *counters*, where a perspex transfers control to itself until some logical condition is met. Thus, the neuron performs the logical part of the perspex instruction: $\mathrm{jump}(z_{11}, t)$. Hence any computer program, written in perspex instructions, is a network of perspex neurons. Conversely, any network of perspex neurons is a computer program. Programs and neurons are interchangeable.

Neurons connect together at *synapses* on the body o of a neuron. Thus, a single neuron has seven synapses where its own dendrites connect to cell bodies, but it can have the use of many more synapses than this. Biological neurons have up to 60 000 synapses. A perspex neuron can achieve this degree of connectivity by passing control to other perspex neurons that read and write data for it. Alternatively, a perspex neuron can use a counter to change itself so that it reads from, or writes to, successive locations of programs space. Perspex neurons are not exactly like biological neurons, but a perspex neuron can emulate any property of a biological one.

The most significant fact about perspex neurons and programs is that they inherit properties from each other. The addition of any kind of structure makes perspex, neural networks more structured, like an animal brain, and makes perspex programs more structured, just like well written, conventional programs. In the chapter *Intelligence* we see how this specific kind of interchangeability, or homomorphism, can lead to the development of highly complex, perspex brains, and efficient, perspex programs, when a perspex robot interacts with the world. Here we concentrate on some elementary properties of perspex neurons.

Perhaps the simplest property of computer programs to note is that logical operations are relatively rare. Typically, nine tenths of what a program does is arithmetic, with just one tenth logic. This means that nine out of ten perspexes are connected to each other in a line, with only one in ten perspexes branching in a new direction. We can, of course, force the line onto an arbitrary curve, but the fundamental fact is that perspex neural networks have the local topology of short, neural fibres.

It is possible to have a highly branching, perspex, neural network with no fibres, but it is an empirical observation that programs that do useful things in the world do not work this way. We experience things happening in fairly long chains of events, before we ascribe any selection of alternative outcomes to them. The more predictable the world is to our minds, the less logic we need to write programs that deal with the world, so the longer the perspex neural fibres are. If a perspex robot were to write its own programs, as we discuss in the chapter *Free Will*, the length of the fibre bundles would provide one, objective, measure of how predictable the robot finds its world.

Perhaps the second-simplest property of computer programs to note is that they use arrays of data. Words, sentences, databases, pictures, sound files, text files, in fact, all files, are stored as arrays. As its name suggests, an *array* is an arrangement of data *arrayed* in rows and columns. This property of having data stored side by side in a program is inherited by perspex neurons. The programs that read a word and copy it to a file turn into fibre bundles. There is one perspex neural fibre for each piece of data, and the fibres lie side by side in exactly the same way as the data. The most striking example of this is a computer picture. An image might be made up of a million pixels, or more, so this many perspex fibres line up in a bundle. This means that computer programs that handle pictures are composed of vast neural fibre bundles, much like the visual pathways in an animal brain. The number and arrangement of fibres in a bundle shows us how data is arranged in a perspex neural network. It can show us the major data pathways and large-scale geometrical arrangements, but it will not always be easy to decide how many fibres are in a bundle or what their arrangement is. In the limit, there is just one fibre in a bundle, so we see nothing more in the bundle than in a single fibre.

The third thing to notice is that programs are divided up into subroutines. A subroutine always protects its data from interference from other subroutines, and, if it is well written, it does one particular job. It is a specialist piece of program that can be called on to do its job anywhere in a larger program in the certain knowledge that no damage will be done by subroutines interfering with each other. Surprisingly, perspex neurons can be organised into functional blocks, or subroutines, that cannot interfere with each other. Perspex neurons inherit this property not from programs, but from spacetime.

We perceive the world we live in to be composed of three spatial dimensions, call them length, breadth, and height, x, y, and z, or anything you like. The important thing is that there are three spatial dimensions. We also perceive the world as changing over time. Mathematicians describe time as another dimension, call it t, that is geometrically the same as the spatial dimensions. Together the three spatial dimensions and the time dimension make up a four dimensional, geometrical spacetime where all dimensions are treated equally. However, physicists usually interpret their calculations in special ways so that they describe what goes on in a 3D spatial world where time flows forward. Such special interpretation is not needed in perspex spacetime, because the forward flow of time is an inherent part of the perspex neuron.

When the perspex neuron performs the matrix multiplication it normalises the resultant matrix so that t_4 is equal either to one or else to zero. The perspex instruction always takes a step along the time axis by the amount t_4 so it either steps by no time, or else by exactly one unit of time. All of the active perspex neurons that are connected in a network with a zero time step appear in 3D space at an instant in time, but when a neuron steps one unit along the time dimension it moves into a new 3D space and can never return, because t_4 can never be negative. These 3D spaces at fixed times provide subroutines. The processing in one 3D space is perfectly protected from interference from the processing in another, because the perspex machine cannot move backwards in time. This issue is explored more fully in the chapter *Time* where a theoretically possible form of time travel is discussed, along with an explanation of what time travel would mean for a perspex machine.

Sadly, we do not know how to build machines that operate everywhere in 4D spacetime, but we do know how to build machines in 3D space that change over time. All we need do is arrange that all of the perspex neurons in one 3D space go into one functional block of neural tissue, with fibres connecting functional blocks that represent the temporal sequence of 3D spaces. It will usually be sensible to arrange these functional blocks close together in 3D space, so that the connecting fibres are short. If we want to, we can physically intermingle functional blocks in one block of 3D space. The functional blocks of neurons will be kept functionally separate by the topology of their interconnections. In the next chapter we see that there is a good reason why we might want to intermingle functional blocks in 3D space and why we might want to arrange neurons in particular 3D shapes. But let us look now at one more, extremely useful, property of programs.

Many programming languages allow us to manipulate matrices. Hence we can implement a model of the perspex machine in a standard computer[1]. However, the details of creating a new matrix and copying parts of matrices into each other are fairly technical. The perspex machine can compute any matrix it needs by multiplying a few standard matrices together and using logical jumps, but this can be an intricate and time consuming task. Some shorthand for reading and writing whole blocks of a matrix speeds up the process. This is what the *access column* provides.

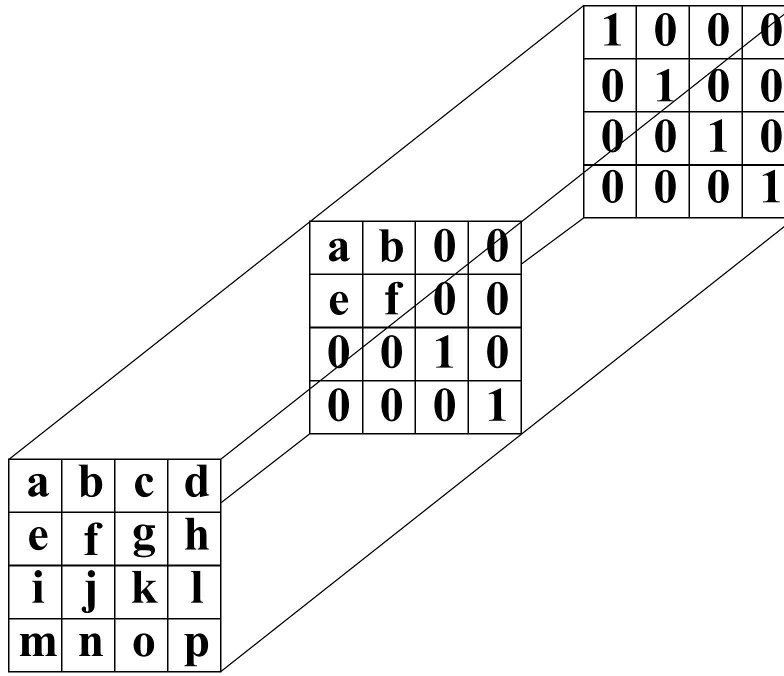


FIGURE 3. **Perspex Access Column Analogous to a Primate Visual Column**

The access column is a column of perspex matrices that has special access properties that control what information can be written into the matrices in the column. The first matrix allow every element to be written to. In Figure 3 it is shown with elements *a* to *p*. The last matrix has the fixed elements of the identity matrix. That is, it has ones on the diagonal and zeros everywhere else. The last matrix and the intermediate matrices can never be written to, but all of the matrices can be read from. The intermediate matrices contain different combinations of the elements from the first and last matrix. In theory we need 2^{16} matrices in the access column to account for all possible combinations of the elements of the first and last matrix. This is 65 336 matrices. This might seem like a lot of matrices to use to speed up perspex programs, but in many practical cases we would not need them all. Alternatively, we could construct a perspex brain with just one access column, but it would be very heavily used, and damage to it would almost certainly devastate the perspex mind.

It is almost certainly a pure accident that visual columns in the primate visual cortex have up to 10 000 neurons[4], which is the same order of magnitude as the number of perspex neurons in an access column. This does, however, give us some reason to believe that perspex brains might contain very roughly the same number of neurons as an equivalent biological brain.

With the aid of an access column, a perspex machine can copy blocks of elements very quickly by writing to the first matrix and reading from whichever matrix is desired. More complicated schemes are also possible. We could create access columns where we can write into the first and last matrices, thereby obtaining mixtures of two arbitrary matrices. We could also allow writing into any matrix such that only the non-identity parts are copied to other matrices. This would allow arbitrary copying of blocks of a matrix. All of these operations would speed up perspex programs considerably, but they do raise a number of delicate mathematical issues and one practical one. There is no efficient way to construct a perspex, access column using only perspexes – if there were, we would have no use for it because we would use this method instead of an access column. If we want efficient access columns we must construct them out of smaller elements than matrices; we must, in fact, do all of the technical work of copying blocks of matrices or, what is more efficient, transcribing blocks using indirection. This is an exercise for programmers.

It is mildly interesting to note that the access columns share some properties with visual columns in the primate, visual cortex. Visual columns are thought to identify the orientation of lines on the retina. Access columns could be used to do this, but they have much more general uses – they speed up all manner of perspex programs.

It would be interesting to know if any other kinds of neural structure would speed up perspex programs. The practical need for fast computation might force us to develop a wide variety of perspex structures to match the many different kinds of biological neurons and neural structures. The demands of living in the world might force perspex robots and animals to evolve similar neural structures.

It is now time to examine the other side of the inheritance, or homomorphism. What properties do perspex programs inherit from brains?

When you look at a light-buff coloured brain, gently pickled in formalin, there is not much about its structure that is obvious. I suppose you might notice that the outer parts of a brain are crinkly. This makes an efficient liquid-cooled heat exchanger, but that is scarcely relevant to a brain's mathematical properties.

One striking thing about a brain is that it is bilaterally symmetrical, as, indeed, are most animal bodies. This might just be nature's way of ensuring that all the parts of an animal join up, but in the chapter *Intelligence* we will see that symmetry plays a profound role in forcing perspex machines to be intelligent. If animal brains work like perspex ones the arrangement of their internal parts should be highly symmetrical. This might show up in the whole brain as a gross bilateral symmetry, but it is not the only way to arrange a symmetrical brain. Perhaps nature's way of making sure that body parts join up is entirely responsible for the bilateral symmetry of an animal brain. If so we should look somewhere else for insights into good ways to make perspex programs.

The most obvious thing about animal brains is that they develop before birth, grow and die to a limited extent during an animal's life, then die completely on an animal's death. Perspex machines are similar.

By default every location of program space contains the halting instruction[1]. This instruction stops a program. A program space that contains no perspexes other than the halting instruction just stops. It is analogous to dying at the instant of birth.

If a perspex machine is to do anything it must have some initial perspex programs laid out as perspex neurons in program space. One of these neurons must be at the origin of Euclidean space because a perspex machine starts by executing the perspex at this location. If the perspex machine ever halts it re-starts at this location. Unfortunately, a perspex machine could kill itself by the simple act of writing the halting instruction into the Euclidean origin. In a serial program there is a significant risk that the perspex machine will do this, or that it will overwrite an early instruction with the halting instruction, but in a parallel program with many starting points it is increasingly unlikely that the machine will overwrite instructions closely connected to all of the starting points. Such a machine cannot kill itself by a simple act of thought, but a complex thought might do the trick. Fortunately, we can build in stronger safeguards.

The initial programs that a perspex machine must have, so that it can do anything at all, are analogous to the development of an animal's brain before birth.

During normal processing, perspex programs write into locations of program space. This is analogous to growing a neuron at each location written into. Unchecked a perspex program would rapidly grow into all of the available space and would start overwriting itself. This would lead to a form of senility and, potentially, death. A program can be checked by taking care only to write into specific locations and to re-use these; but a program that has its own creativity or free will cannot be constrained in this way. It is free to bring about its own senility and death, as discussed in the chapters *Free Will* and *Spirituality*.

A perspex program can free computer memory by killing unwanted neurons by writing the halting instruction into the location of a neuron. This destroys the neuron at that location and consumes no computer memory of its own[1,2]. Thus, a perspex machine can re-use its own memory. A perspex machine that has its initial, birth programs written into read-only memory can be protected to an arbitrary degree from killing itself by an act of thought, but a perspex machine that is not entirely composed of its birth program always runs some risk of senility and death.

Thus, the perspex neuron provides not only the computation essential to a robotic mind, but also lays the foundations for growth and death. Perspex robots with creativity and free will will be mortal and risk madness, just like us.

Of course, perspex neurons are just one way of seeing the perspex instruction. We saw in the chapter *Perspex Matrix* that a perspex can be seen as a motion. In the chapter *Perspex Simplex* we saw that a perspex can be seen as a shape. This has a profound influence on the role of language in a perspex mind, as we discuss in the next chapter. It also has practical consequences. Shapes and motions describe all of the physical universe, so a perspex machine can be embodied in any physical thing. It also means that any physical science might hold useful lessons for the design of perspex machines. We see in the chapter *Perspex Instruction* that the perspex machine can solve all Turing-computable problems and more, so it can, presumably, solve all problems that the human mind can solve. Thus any science, including the abstract sciences of mathematics and philosophy, might hold useful lessons for designing perspex machines. The perspex machine speaks to the whole of the human condition, as we discuss in subsequent chapters.

Questions

1. Can the shapes and motions encoded in a perspex neural network be displayed by a computer animation so that we can see what the network itself is visualising?
2. Which neurophysiological techniques can be applied to computer graphic animations of perspex, neural networks? Can we see neural networks grow and die? Can we trace afferent, efferent, and transferent pathways? Can we highlight fibre bundles and functional blocks? Can we highlight access columns, counters, or other neural structures?
3. What kinds of access column are useful, and which have the most succinct mathematical interpretation as a product of elementary matrices?
4. Water has a much higher thermal conductivity than air so marine mammals are in less need of biological structures to cool their brains than terrestrial mammals. Do marine mammals have less crinkly brains than terrestrial mammals? How crinkly are reptile brains?

Further Reading

1. **Anderson, J.A.D.W.** "Perspex Machine" in *Vision Geometry X1*, Longin Jan Lateki, David M. Mount, Angela Y. Wu, Editors, *Proceedings of SPIE* Vol. 4794, 10-21, (2002).
2. **Anderson, J.A.D.W.** "Robot Free Will" in *ECAI 2002 Proceedings of the 15th European Conference on Artificial Intelligence,* Lyon, France, ed. F. van Harmelan, pp. 559-563, (2002).
3. **Churchland, P.M.** *The Engine of Reason, the Seat of the Soul: a Philosophical Journey into the Brain* MIT Press, (1996). First printed 1995.
4. **Popper, K.R. & Eccles, J.C.** *The Self and its Brain - an argument for Interactionism* Routledge, (1995). First published 1977.
5. **Turing, A.M.** *Computing Machinery and Intelligence* Mind, volume LIX, number 236, pages 433–460, (1950).

CHAPTER 6

Beyond Language

Introduction

"That deaf, dumb and blind kid sure plays a mean pin ball," according to the *Pin Ball Wizard*, by the rock band, *The Who*. This might be a good rock song, but it is terrible science. If the kid is blind he has no way of sensing where the ball is and will soon fail due to the chaotic nature of pin ball. But the song does raise an interesting question. How is it that someone who is deaf, dumb, blind, and illiterate can dress, eat, organise themselves, and carry out useful work? How do such people think? Do they have an internal language that allows them to think? When they learn to communicate with other people via touch, do they do so by relating touch to their internal language? How is it that for most people an internal language can relate to such diverse things as spoken vibrations of the air, hand written signs on paper, and the striking and caress of fingers on the hand? And what of non-human, sentient beings? How do cats think? Do cats have an internal language? How might robots think? Do robots have to have an internal language?

There are many possible answers to these questions, but we now look at one theoretical answer. Thinking is not done in any language, but in visualisation. Visualisation is more powerful than language, but if it is sufficiently restricted it can become language. Visualisation is tied to spacetime by a causal chain of perspex actions so it relates causally to spoken vibrations of the air, hand written signs on paper, and the striking and caress of fingers on the hand, or to anything whatsoever. The form and meaning of words and linguistic structures is not arbitrary, it is tied to the perspex structure of a brain embedded in a body. The fact that we have brains and bodies that operate at a similar scale in spacetime to those of cats means we can communicate with each other. Not because we share a language, but because spacetime constrains our brains and bodies in similar ways. We could also communicate with a robot or an extraterrestrial being at our scale of spacetime. Such is the generality of the perspex.

Turing and Perspex Machines

Alan Turing worked out a most amazing story or thesis[14,15,16]. He said that he did not know how people think, but that it will be commonly agreed that people cannot keep track of very large problems in their minds, but write down, on paper, intermediate notes and instructions on what to do next. It is then, he supposed, a physical fact that there is some smallest size that a mark on paper, or on anything, can be. Consequently, no matter how many marks it might be possible to make in the universe, they are countable at this smallest scale. Being a mathematician, Turing then decided that he would use integers to represent the marks, or *symbols,* as he called them. He then defined a theoretical machine with two parts. The first part of the Turing machine, as it is now called, is an internal memory with a finitely long program in it. The second part of the machine is a data tape that has a finite number of symbols on it at the start of a computation, but which is potentially infinitely long so that arbitrarily many symbols can be written onto it. The Turing machine can do four things: it can read a symbol from the tape, write a symbol onto the tape, move the tape forward by one symbol, or move the tape backward by one symbol. That is all is can do by way of actions following its internal program, but the astonishing thing is that this machine appears to be able to compute anything that can be computed by any other mathematical definition of computability. The thesis that human beings and physical machines can compute no more than a Turing machine is called the *Turing thesis* or, sometimes, the *Church-Turing thesis* – in joint honour of Alonzo Church who also worked out a mathematical theory of computability. It appears that the Turing thesis is correct and, today, the hypothesis that all physical computations are equivalent to Turing machine computations is widely accepted; though the philosophical consequences of the thesis are still debated. It remains possible, however, to define *theoretical* computers that do more than a Turing machine. It remains an open question whether any super-Turing machines have already been built, or can be built[6]. The Turing thesis does not define what computability is, it defines a Turing machine, which is hypothesised to explain all physical computation. The definition is right by definition, but the hypothesis might be wrong.

Now let us try to linguistically translate the Turing machine into geometry. We might imagine a train running on tracks that can move forward or backward one sleeper at a time and which can pick up or put down a symbol on each sleeper. We could imagine reducing the train to a point, so that it can be part of a geometry. We could imagine geometrically translating the train along the track by changing its origin, making geometrical translation part of the geometry, but there is a problem. We are interested in the kinds of geometry where space is continuous, not in the kinds of geometry where space exists in discrete steps that can be numbered one after another. It would appear that we cannot linguistically translate the Turing machine into a continuous geometry that describes the space we live in. And there is another problem: we do not live in a one dimensional world where the only possible direction of movement is forward and backward.

The ability to take discrete steps that can be numbered one after another is fundamental to the Turing machine and has profound consequences for the kinds of number it can compute. A Turing machine can generate all integers one after another by using a counter, or, more generally, all rational numbers in an arbitrary order by using a counter and some encoding of rational numbers as integer symbols. If a Turing machine has a way of recognising that a number has some desired property it can stop when it gets to this number and can print it onto the tape as an answer. Gödel's celebrated undecidability proofs[10] show that there are always some properties of numbers that any given Turing machine cannot recognise. There is always something a Turing machine cannot compute. Gödel's proofs hinge on the fact that all symbols, like Turing's symbols, can be encoded by integers. But if we succeed in linguistically translating the Turing machine into a geometry where space is continuous, Gödel's proofs will not apply. If we succeed, we will produce a theoretical machine with more computational power than a Turing machine. If we can build this super-Turing machine in our physical universe we will have proved that the Turing thesis is false.

Turing came up with another kind of computability. He said that if a Turing machine can compute that some estimate of a number is too large, or too small, then it can take steps toward the correct answer and can print it onto the tape to ever greater accuracy, even if it can never print out the exact answer. Thus a Turing machine is able to compute some of the irrational numbers such as $\sqrt{2}$, π, and e, but not all of them. All of the irrational numbers are interwoven with all of the rational numbers to form the continuum of real numbers on the number line. However, Turing also proved that there are some real numbers that a Turing machine can never compute, because it cannot tell how close any number is to the true answer. All decimal numbers with an infinite number of random digits have this property, so they are Turing-incomputable numbers. If we succeed in translating the Turing machine into a geometry where space is continuous we will, theoretically, be able to compute Turing-incomputable numbers. If we can build such a physical machine we will have proved, again, that the Turing thesis is false.

Turing has one more, rather subtle, clue for us. He did not use all of the integers as symbols just zero and the positive integers. This does not restrict the kinds of numbers a Turing machine can describe. For example, a Turing machine can describe negative integers by using a numerical code as a minus sign to indicate that an integer is negative. But the use of non-negative integers as symbols does tell us something profound about how a Turing machine divides up the space of numbers. In general, a Turing machine uses a counter to start from zero and count up to whatever number it needs. The Turing machine can only count up from zero, it has no natural way to count down, though it can arrange to do this using some rather convoluted encodings of counting up how many steps down it needs to count. Turing machines use some pretty bizarre programs, but, fortunately for students, there are some definitions of computation that are much closer to normal programming. If you are a student, and want to take the easy and productive route to understanding Turing machines, study Unlimited Register Machines[8] instead. But, for our purposes, we have already discussed all the properties of Turing machines that we need in order to understand the perspex machine.

In the space we live in we have no problem walking forwards and backwards, or counting up and down so Turing's only-counting-up is not going to linguistically translate into a geometry of the space we live in, but it might linguistically translate into a geometry of time. In our ordinary experience of the world, time always moves forward. We can live our lives forward in time, but we cannot live backwards in time so only-counting-up might well linguistically translate into a geometry of time. In fact, a notion of time is already built into Turing machines. Turing defined that each operation of a Turing machine takes exactly one unit of time, so this defines Turing time. In Turing time operations take up a contiguous block of elapsed time and time never moves backwards.

Drawing all of this together, if we want to linguistically translate Turing computation into a geometry of the space we live in, we can see: we have to be able to count up and down in space; we will accept only-counting-up in time; we need more than one dimension of space; and we want to be able to compute in the continuum, not just in integer or rational numbered points in space. So let us move a little further toward our goal, let us see what properties a geometry of two spatial dimensions and one temporal dimension has.

There are many ways we could lay out such a geometry, but here we use a 2D perspex like the 3D perspexes in the chapters *Perspex Matrix* and *Perspex Instruction*. A 2D perspex is a 3×3 matrix of homogeneous co-ordinates. Let us call its column vectors x, y, and t. We agree to read a 2D perspex from program space at the location x and write it to a special location of space that is an accumulator. On the first occasion, and subsequent odd-numbered occasions, when a perspex is written into the accumulator the accumulator is cleared of any matrices that may have been in it. On the second, and subsequent even-numbered occasions, when a perspex is read from some x location into the accumulator, the two matrices in the accumulator are multiplied together. The result is normalised and written into the location y. The top-left element of the result, that is $\vec{y}_{11}$, is examined. If it is greater than or equal to zero control is passed along the x-axis by the amount t_1. Otherwise, if $\vec{y}_{11}$ is less than zero control is passed along the y-axis by the amount t_2. In every case control is passed along the t-axis by the amount t_3. There is just one safe place to put the accumulator, where it cannot be overwritten by y, and that is at the origin of homogeneous space: $(0, 0, 0)$.

A slight modification of the proof published in[2] then shows that the 2D perspex machine manifests the same concept of time as the 3D perspex machine – we will discuss this concept in the chapter *Time* – and can do everything that the 3D perspex machine can do, all be it in a more long-winded way. For example, the 2D perspex machine has no natural way of testing if a number is equal to zero. It must, instead, check to see if both the number and its negative are greater than or equal to zero. This is the way that French mathematicians talk about numbers. They say that the continuum of the number line is divided up into two parts: the negative numbers and the positive numbers. According to the French, zero is a positive number. But, so far as I know, mathematicians from every other linguistic community in the world say that the continuum of numbers is divided up into three parts: the negative numbers, zero, and the positive numbers. This view of numbers is formalised in the *triality axiom* of number, which mathematicians from every linguistic community are willing to work with. But the harsh fact remains that, when it comes to the concept of number inherent in a perspex, 2D perspexes are French and 3D perspexes are not French.

It is astonishing to think that national character might develop in a robot as a consequence of the mathematical assumptions built into a perspex, but this is a real risk. If the perspex suits one kind of human language better than another then robots in the well matched linguistic community will acquire its language and culture more quickly than in a community using a badly matched language. There is also a risk that robots with similar bodies or brains will divide up into robot nations. Racism need not be an exclusively human phenomenon.

There is, however, a much stronger consequence of the choice to use 2D or 3D perspexes. This can be understood by an appeal to projective geometry[7], or by waving ones arms about.

If you have two arms each with a pointing finger, a table, and the power of voluntary movement, then clear the table and kneel down, or otherwise arrange your body so that your left arm lies flat on the table. Imagine an infinitely long, straight line that lies on the table top and passes through your left shoulder and your left pointing finger. Call this line the, “arm line.” Now put your right hand flat on the table. Imagine an infinitely long, straight line that lies on the table top and passes along the axis of your right pointing finger. Call this line the, “finger line.” Now point with your right, pointing finger toward your left shoulder. Let us define that your left shoulder is at the distance zero along the arm line. Now point with your right pointing finger to the tip of your left pointing finger. Let us define that this is at the distance plus one along the arm line. Now point to this distance on the arm line behind your left shoulder. This is the distance minus one on the arm line. Keeping your left arm flat on the table, and your right hand flat on the table so that the two lines are in the 2D plane of the table top, answer the question, “Can I point at anywhere, with my right pointing finger, so that the finger line does not meet the arm line?”

If you point with your right pointing finger so that the finger line is parallel to the arm line then, according to Euclidean geometry, you are *not* pointing at a point on the arm line, you are pointing at a point infinitely far away in the direction your pointing finger is pointing. But, according to projective geometry, parallel lines do meet at a single point at infinity. It is like two, infinitely long, parallel, railway tracks meeting at the horizon. Notice that these tracks meet at the horizon both in front of you and behind you. According to projective geometry a line points in both directions along itself, so that plus infinity and minus infinity are the same place. Therefore, when the finger line is parallel to the arm line they meet at just one point at infinity. Now imagine rotating your right hand in a full circle on the table top. At every orientation the finger line intersects the arm line somewhere. This is a rather neat result. It means that every question you can ask about where lines meet can be answered. But this neatness comes at a price, you are locked into a world with two spatial dimensions. If you doubt this, then answer the question, “Can infinitely long, straight lines fail to meet or be parallel?”

I teach computer graphics to many students. In my experience about one in forty cannot visualise 3D space and so cannot answer the question just put. Perhaps you are one of these people? I have heard of two people, other than myself, who can visualise 4D spaces, and higher. However, such unworldly visualisation does seem to be rare. I have never knowingly met such another person. It seems that most people visualise the 3D world in which we live. This is the world we want perspex robots to see so that they can share the common human perception of the world and can understand our languages.

With your left arm flat on the table, lift your right hand off the table. Now point with your right pointing finger somewhere that is not on the arm line. This is easy. Just make the finger line and the arm line skew. Almost every volume of space that you can point at is not on the arm line. In these places projective geometry cannot say what distance along the arm line you are pointing, nor can any mathematics that accepts the triality axiom. This has serious consequences for a robot.

Imagine you are a robot sitting in a room. Your eye alights on the straight line between one wall and the ceiling and then on the line between an adjacent wall and the ceiling. You happen to ask yourself where the two lines meet and correctly predict the position of the corner of the room. Now imagine that you are walking outside. You happen to see one line at the top of a wall, and then another line at the top of a different wall. You happen to ask yourself where the two lines meet; instantly your thinking is disrupted, or you die. Unluckily for you, the two lines happened to be skew to each other, as happens most of the time in the world. There is no way that your mathematical brain can answer the question of where the two skew lines meet, so it went into an error state, seized up, or died.

Now imagine that some damn-fool human asks you, "Which weighs more heavily on you: a ton of bricks or the fear of failure?" There is no way to calculate the two weights involved, so your mind is again disrupted, or you die again. (It is rather easy to resurrect a robot, unless it is seriously senile.)

Life would be so much easier for a robot if it could have a number like *nullity*[3] that does not lie on the number line. Then, in the first place, the robot computes that the two wall-lines meet at nullity and correctly predicts that the tops of the two walls do not meet anywhere in the space of the world. In the second place, it computes that the axis on which weight-of-bricks is measured meets the axis on which weight-of-failure is measured at nullity. In other words, these two axes do not intersect anywhere in the robot's mental space so it knows that weight-of-bricks and weight-of-failure are incommensurable. A little thought then convinces the robot that most of the concepts in its mind are incommensurable. Enriched by this thought the robot answers, with a laugh, "You are not heavy. You are my brother."

This story, and the argument it illustrates, throws out a very minor challenge to mathematics. If we want mathematics to describe 3D projective geometry in the simplest way, then we must extend the triality axiom to a quadrality axiom by embracing the number[3], or point[1], at nullity.

Thus we arrive at the 3D perspex as being the most natural way to combine the Turing machine with the geometry of the space we live in and see. But there are at least three ways in which a theoretical, perspex machine can do more than a Turing machine.

Firstly, consider the perspex machine which operates by light shining through a pin-hole. The light may come from any point on a line so it may, theoretically, represent any number on the number line, including Turing-incomputable numbers. If the pin-hole perspex machine can operate exactly, so that it can tell if a number is exactly zero, then it can test any number for equality with a Turing-incomputable number. That is, the exact perspex machine can do more than a Turing machine. However, it seems rather unlikely that any physical machine can be exact, and it remains an open question of physics whether light can take up a Turing-incomputable position on a line[12,17].

Secondly, whilst some Turing-incomputable numbers cannot be written onto a data tape, because they cannot be written as a sequence of digits, some can be written as an infinitely long sequence of digits. These latter, *enscribable,* Turing-incomputable numbers can be given, by fiat, to a Turing machine. The Turing machine cannot compute anything with the whole of a given, enscribable, Turing-incomputable number, because it has infinitely many digits, but it can truncate it to a Turing-computable, rational number and operate on that to an accuracy limited by the truncation. That is, the Turing machine can operate to limited accuracy on a given, enscribable, Turing-incomputable number. The Turing machine can also engage in the never ending task of carrying out a point-wise, computable, transformation of a given, enscribable, Turing-incomputable number into a different enscribable, Turing-incomputable number, but it cannot carry out any incomputable transformation to produce a radically novel, Turing-incomputable number. However, if the physics of light allows the perspex machine to be given a Turing-incomputable number[12,17], it can operate on the whole of this number to an accuracy truncated by physical limits. To this extent, the limited accuracy, pin-hole, perspex machine is like the Turing machine. But if light can take up any possible Turing-incomputable position then the pin-holes can take up any possible Turing-incomputable positions relative to the light and can, thereby, perform Turing-incomputable operations. That is, the limited accuracy, pin-hole perspex machine can carry out incomputable operations and can compute radically novel, Turing-incomputable numbers, whether or not they are enscribable. Thus, the perspex machine passes beyond any language.

Thirdly, a variant of the perspex machine might exploit not pin-holes, but lenses with a variable optical density. In this case computation does not occur at a finite number of pin-holes, but in the body of the lens. It remains an open question of physics whether a lens supports a continuum of positions of light or whether only a countable number of such positions occur. If a continuous perspex machine can be built then it can simultaneously undertake an infinite number of calculations and can complete them all in finite time. This is more than a Turing machine can do.

My best guess is that, of the three kinds of device considered above, only the inexact pin-hole perspex machine will make a practical super-Turing machine[6]. But for our purposes it is enough to note that the theoretical perspex machine is super-Turing.

Language is a symbolic system and can be described by a Turing machine. Turing-incomputable sentences can be described in the same way that Turing-incomputable algorithms can be described: a Turing machine can have them on its data tape, even if it cannot do anything useful with them. No matter what we express in language a perspex machine can express this much, and more. Language is contained within perspex visualisation. This gives us a new way of *looking* at language, which can only be of benefit to linguistics.

Philosophers have a more pointed challenge. Logic is a symbolic system so it too can be described by a Turing machine. But the perspex machine can do everything that a Turing machine does, and more, so the perspex machine can do everything that logic does, and more. For example, to say that some predicate is logically necessary does not say that it is necessary. A perspex machine might have some other way of dealing with the predicate. If philosophers allow this possibility then they are going to have to be much more careful about their claims for logic and metalogic (the logic of logics). See, for example, the chapter *Spirituality*.

Meaning

Some philosophers worry over the supposed fact that the meaning of words is arbitrary. They imagine that a language can be made up of any symbols whatsoever and that any particular meaning can be attached to any symbol. From a materialistic point of view this is plainly wrong. All abstract things are ideas described by perspex neurons in a perspex brain, biological neurons in an animal brain, or some physical kind of thing in any physical kind of brain. Abstract things do not have any existence apart from their physical basis. The physical basis constrains the symbols in a language; constrains the meanings in a brain; and constrains which specific meanings are attached to which, specific, symbols in a brain. Minds do not have any existence apart from their physical basis. But see the chapter, *Spirituality*, for a discussion of materialism which is consistent with a deity, heaven, and life after death.

Some philosophers argue that it is logically possible that the physical universe does not exist, so that everything is purely abstract. There is a standard answer to them. They are logically correct, but if the physical universe is, say, a dream in the mind of God then the dream is what scientists mean by the “physical universe,” and the mind of God is the ultimate "physical" reality, even if this reality is not accessible to us. It makes no difference to us whether the "physical universe" is material or purely abstract. So, let us continue to talk about "the physical universe" regardless of its ultimate nature. In particular, let us talk about the perspex to the extent that it can describe the whole of the physical universe.

Start by imagining how we might build a robot that has the ability to learn language in a human culture. In order for the robot to become a member of our culture and share the assumptions, advantages, and responsibilities of society we choose to make it as much like us as we can. We choose to make it an android in our own image. We start by making a mechanical body in the shape of a human body. If we lack the motive power or engineering skills to construct some part of its body then the android will be a handicapped member of our society. It will certainly be mentally handicapped as we struggle to design and construct a perspex brain capable of developing a mind like ours.

Having constructed a robot body we then construct a birth program for it and write this into read only memory in the android's perspex brain so that it cannot kill itself by a simple act of thought. Some of the technicalities of doing this are discussed in the chapter *Perspex Neuron*. The way we write the program is constrained by the human-like body and our desire to make the android's mind as much like ours as possible. All of the joints in the android's body will be hinged or rotational joints, like ours. The way these joints move can be described by a perspex matrix that encodes rotation, as discussed in the four introductory chapters. We choose to lay these perspexes out in the android's brain in a topographical map portraying the shape of its body. We arrange electromechanical controllers in its body so that writing a perspex into the location describing each joint causes the joint to move to that position. Thus we provide the robot with a perspex motor cortex analogous to our own motor cortex[11]. We call the perspexes that control the position of joints *motor* perspexes. We also arrange to put rotational sensors, *proprioceptors*, in the joints that record the position of the joint and write this position into a topographic map of perspexes. If we want the android to have fast control of its limbs, so that it can check that it is moving its limbs as instructed, then we place the motor and sensory perspexes close together in the android's motor cortex. This is a case where there is an advantage in intermingling different functional blocks of perspex neurons in one geometrical block. If we want to provide finer control of motion then we will set up a topographic map of differences in instructed joint position, and sensed velocity of joints so that we can control the first, or higher, differentials of limb position. That is, we provide the android with *kineceptors* analogous to our own. When attempting to control high differentials it is extremely important to keep the data paths between the sensory and motor parts of the motor cortex as short as possible. Thus, the need for fast and accurate reflexes will shape the geometry of the perspex motor cortex.

Space rockets use sensorimotor controllers up to the fifth differential which provides far more control than in any animal body. This indicates that robots might, eventually, have far finer bodily control than animals.

No doubt there is a great deal more that the birth program should contain to give the android abilities analogous to a human child at birth and to provide for similar mental development. We see what some of this might be as we continue our imaginary exercise of allowing the android to acquire the bodily and sensory skills sufficient to learn language.

We suppose that the android has been programmed to move about and calibrate its body and brain so that the motor perspexes accurately instruct motion of its limbs. Furthermore, we suppose that the android is programmed to reach out for, and to explore, any new object in its environment, and to look toward the source of any sudden sound. We imagine that the android has had long exposure to a human "parent" since birth, so that the parent is no longer a novel object. Now suppose that the parent brings in an unbreakable, plastic cup and says, "cup" the android, following its program, looks toward the source of the sound and, coincidentally, looks at the cup. When it reaches for the cup the parent repeats the word "cup." If the android looks away the parent puts the cup in the android's line of sight and says, "cup." If the android does not take hold of the cup the parent puts the cup in the androids hands and says, "cup." What happens next?

The android's fingers are in contact with the cup so the motor perspexes which control the position of the android's arms, hands, and fingers record the position of the contact points, as do the sensory perspexes. If the robot manipulates the cup in its hands, and explores its surface, the sensory and motor perspexes each trace out geometrical models of the shape of the cup in some co-ordinate frame peculiar to the geometry of the android's arms. Providing the android keeps a memory of its actions and sensations it will grow at least two geometrical perspex models of the shape of the cup: one motor model and one sensory model. Each of these models records, at a minimum, the contact points of the fingers on the cup.

In the chapter *Visual Consciousness* we see that in order to obtain consciousness, as defined in that chapter, we need to relate geometrical models to each other in a way that allows a brain to pass from points on one of the models to points on the other, and back again. That is, we need the perspex brain to grow a bi-directional relationship between the sensory and motor perspexes describing the cup. One of the easiest ways to do this is to grow perspexes that jump from one model to the other. These intermediate perspexes relate the sensory and motor parts of the perspex motor cortex to each other and can mediate a primitive language internal to the android.

This language might, for example, be categorised as having primitive nouns – all of them positions of the arms, hands, and fingers – and primitive verbs – all of them motions of the arms, hands, and fingers. Sentences might have a primitive grammar of repeating couplets of a noun and a verb that describe the current contact point on the cup and the verb required to move the fingers to the next contact point. Each time the android handles the cup it learns a new sensorimotor sentence describing it. Thus the android's internal language is related by a causal chain to bodily motions, sensations, and things in the world.

These relationships are primitive, but they need not be expressed in language. They can be expressed in perspexes that describe positions and motions directly. It is then a contingent fact of physics whether these positions and motions are all described by computable numbers, and hence by Turing symbols, or by incomputable numbers. In the latter case there is no possible, symbolic, language that can exactly express the content of the relationships encoded in the android's brain. It can feel and do things in the continuum that are inexpressible in words. For example, the android can know how to do things, such as holding a cup or driving a train, that can never be put entirely into words. This knowing-how forms part of Chomsky's deep structure of language[5]. It is the meaning behind the facade of words.

But there is no dichotomy here. Perspexes can hold computable or incomputable numbers, they can be symbols or not. There is no necessary difference to a perspex brain between expressing thoughts in language and expressing them in the continuum. Thoughts in the continuum are visualisations and, when restricted to computable numbers, are expressible in, or simply are, symbolic sentences.

There is, however, a contingent reason why we might want to introduce a physical difference between symbols and the continuum. In many uses of language one wants to make exactly the same reference again and again, despite the presumed fact that physics does not allow an exact repetition of worldly circumstances, including exactly repeatable computation in a continuous, perspex machine. In sophisticated uses of language, such as in algebra, we might want "x" to refer to the same thing regardless of how "x" is written on each occasion. This can be done by having each of the sensorimotor perspexes jump to a topographical map, or manifold of the possible visible shapes or writing actions of "x," but having each of these perspexes jump to a single location, being the unique contents of "x." Thus the jump parts of a perspex are all the same in the neighbourhood of a sensorimotor continuum that is a word. Words are flat neighbourhoods in an otherwise continuously varying perspex space. Flat neighbourhoods also buy some resistance to error. If we copy a perspex into a neighbourhood of program space then any point in that neighbourhood is an exact copy of the perspex and performs exactly the same instruction. Thus flat neighbourhoods provide a means of introducing digital symbols into a continuum.

On this account, words can be represented by unique things in a perspex brain, and can be associated with unique meanings. But the words and the meanings are grown in the brain according to causal relationships. The words, meanings, and their connections are not arbitrary they are determined by the physics of spacetime. We take up this aspect of determinism in the chapter *Free Will*.

But there is another way to look at words. Words are not symbols that contain meanings they are the flat parts that separate regions of curvature on a perspex manifold. The manifold is a well behaved surface, like the landscapes we are used to seeing in the world. Meaning does not reside on the flat paths dividing up the landscape, but in the rolling hills. Meaning resides in the curved parts of the manifold because a point on the curve can be selected so that it closely matches the value of any continuously varying quantity in the world. It is this space between words that relates to the world. When a parent says, “squeeze the cup,” or “drive the train faster,” the words in the sentences are boundary markers in a manifold of increasingly tight hand grips and increasingly greater train speeds. Seen this way, words do not contain meanings, instead they operate by moving the language user to a different region of mental space. They signpost the way into the landscape of the deep structure of language that is written in perspexes.

Of course, this is not a dichotomy perspex machines can use words both ways and can use the whole continuum without regard to language.

So far, our imaginary android has learnt some internal, sensorimotor sentences to do with a cup. It has the opportunity to learn at least one sentence every time it picks up and drops the cup. But is there another way for the android to acquire language?

In this section we approached language from the supposed properties of visual consciousness. This might be illuminating, but it will not necessarily answer the question of how a blind child can learn language. However, we can understand language by considering the properties of a single perspex irrespective of its role in vision. Language is a manifest property of the perspex.

Manifestation

Some philosophers and scientists follow the paradigm of *emergence*[11]. They accept that there is no fundamental explanation of intelligence, consciousness or whatever, but suppose that these things emerge from a biological or artificial system that is sufficiently complex. Explanation is found in the specific complexities of the brain or computer they are studying. The more complex the system, the more it exhibits the emergent qualities. Such people hold out the distant possibility that after many specific explanations have been found it might be possible to synthesise them all into an over-arching theory of intelligence, consciousness or whatever.

Others follow what I call the paradigm of *manifestation*. They hold that there is a fundamental explanation of intelligence, consciousness, and everything that is inherent in very simple things. Explanation is to be found at this very simple level and becomes more manifest as more complex systems are considered. For example, I claim that language is inherent in the perspex and becomes manifest in more complex arrangements of perspexes until it is, pardon the pun, manifestly obvious that perspexes have language, consciousness or whatever. I go so far as to claim that the perspex can describe everything in the universe; and can be a mind that has language, consciousness, feeling, free will, and so on. I do not do this out of some grandiose leap of the imagination, but because I designed these things into the perspex. I start with the theory and work out to the experiments.

Seen as explanations there is no testable difference between emergence and manifestation. Each explains that increasing complexity brings with it increasing intelligence, consciousness, and so on. The explanations of any particular example are the same, even at the most primitive level. There is no difference between these paradigms considered as explanations. But paradigms are much more than explanations; they are ways of working. The two paradigms have different psychological effects on philosophers and scientists: emergence appeals to empiricists because it gives a reason to hope that a blind search of many cases will lead to understanding, without having to engage in the overwhelmingly difficult task of hypothesising an explanation of everything before doing the experiments; conversely, manifestation appeals to theorists because it gives hope that a simple explanation of everything can be found before engaging in the overwhelmingly difficult task of conducting sophisticated experiments to test everything. Each paradigm is a psychological help to the philosophers and scientists who adopt it. Because the explanations are the same at every level they translate into each others terms. Empiricists and theorists work together, even if they do not have much time for each other.

It will not surprise you to learn that whilst I use the paradigm of manifestation, I also use the paradigm of emergence. I know how to operate as scientist in several different ways, and use whichever seems to me to be best matched to the problem. Of course, I do not always use a paradigm. Sometimes I just follow my own visualisation of a problem and think up ways of working that will test the idea. Much of what I do is not science. This annoys the Government appointed bureaucrats who seek to measure the science I am employed to do; but if they will use defective paradigms they must expect to be frustrated and to make dreadful mistakes.

In the next section I show how even a perfect symbolic system makes mistakes and is forced to change its paradigms, simply as a consequence of using language. But now, I concentrate on showing how language is manifest in the perspex. Outside the confines of this book, I hope to do something about the bureaucrats.

In the chapter *Perspex Instruction* I show how the perspex can be the instruction: $\vec{x}\vec{y} \to \vec{z}$; $\text{jump}(z_{11}, t)$. I explain this instruction in terms of multiplying together two matrices $\vec{x}$ and $\vec{y}$ to produce the normalised product $\vec{z}$. Control then jumps to a new position depending on the top-left element, z_{11} , of the resultant matrix and the vector t . If these matrices, including the vector t , contain only Turing-computable numbers then the instruction deals with linguistic symbols, but if all of the numbers are Turing-incomputable numbers then it does not deal with linguistic symbols. It seems that linguistic symbols are, at best, a contingent, or emergent property of the perspex. Furthermore, that grammar is emergent. This is all true, within the paradigm of emergence, but there is another way of looking at things.

Look at the first part of the instruction $\vec{x}\vec{y} \to \vec{z}$. What does it look like? Seen one way it is a relationship between three perspexes $\vec{x}$, $\vec{y}$, and $\vec{z}$. It is a re-write rule, or function, in which $\vec{x}$ and $\vec{y}$ on the left of the arrow are re-written as $\vec{z}$. Re-write rules, controlled by a conditional jump, can do everything that a Turing machine can do, including everything that language can do. But what do $\vec{x}$, $\vec{y}$, and $\vec{z}$ look like? Seen one way they are simplexes. Providing the simplexes are not so singular that they are just a point, they contain a neighbourhood of space. This neighbourhood might be a 1D line, 2D surface, 3D volume, or 4D hypervolume, but, whatever it is, it contains a segment of the number line. Any segment of the number line can be encoded so that it contains the whole of the number line, so any segment can contain all of the Turing-computable symbols, suitably encoded. The perspex contains all symbols and all sentences in every Turing-computable language, and more. But can we pick out one symbol or sentence from this infinitude? Of course we can: we stretch each edge of a perspex onto the whole of the number line and use this perspex as the program space so that all perspexes read, write, and jump inside this one. This can involve an infinite regress, but not a harmful one. It is as if the entire universe of perspexes is contained in the perspex shown on the cover of this book. When we think about a perspex, so as to put it into operation, we enter a deeper level of this universe, and so on, until we stop putting perspexes into operation, or until they converge to a limit. In the case that they do not converge we can kill off the infinite regress by having the bottom level perspexes jump to the point at nullity.

Similarly, if we want a non-symbolic function we jump to a Turing-incomputable point inside the perspex. A single perspex, that is at least a line, inherently contains all possible languages, and a great many non-linguistic things as well. These inherent properties become more and more manifest as the complexity of a perspex brain increases. It is a contingent fact whether we select linguistic or super-linguistic things from inside the perspex, but the perspex contains them all.

It is a strange thought that the whole of this book, and of every book, is contained in the figure on the cover of this one; such is the generality of the perspex.

There is a more general way of looking at the perspex. A perspex describes the position of an object, in fact it gives the whole of a co-ordinate frame centred on an object. The matrix multiplication in the perspex instruction $\vec{x}\vec{y} \rightarrow \vec{z}$ allows objects to move, and to select the next movement by performing a jump. I hypothesise that the perspex can describe the whole of the universe in this way. If I am wrong in this, then there is some part of the universe that cannot be described by a perspex, so we are free to chose this part, property, phenomenon, or whatever it is, and use it to construct a computer more powerful than the perspex machine. The perspex machine is just an initial hypothesis about how to describe the whole of the universe, and how to construct a robot that is intelligent, has consciousness, feeling, free will, and any mental phenomenon whatsoever. No doubt there are errors in this hypothesis. No doubt the hypothesis can be improved by both theoretical development and empirical testing. That errors are inherent in the perspex hypothesis, and the amount of work that is needed to correct them so as to bring about a paradigm shift, is explained, in the next section.

Paradigms and Ockham's Razor

Scientists change their paradigms or theories when the need arises, but why does the need arise? Is there any conceivable state of the universe in which a scientist need never change a paradigm? How is it that even computers that operate perfectly, still make errors? Is there any conceivable state of the universe in which a computer would not make an error? Why does trying to improve something that works well, nearly always make it worse? Is there any conceivable way in which corrections would always improve things? Why do politicians so often get things wrong? Is there any conceivable state of the universe in which a politician can always get things right? The answer to all of these questions can be seen by an application of number theory, or by cutting a walnut cake.

Imagine that you like walnuts, but are not particularly fond of sponge or walnut fondant. When offered a walnut cake you always try to cut a slice that contains some walnut, but which has the least possible amount of sponge and walnut fondant. In comes Mrs. Perspex, she has baked a walnut cake, but tells you that, due to a shortage of walnuts, there is only one walnut in the cake and this is a very small walnut; in fact it is just a geometrical point. She calls the walnut a "kernel of truth" and says she will help you find it. If you cut the cake from the centre outward she will tell you if you have cut into the walnut, or, if you miss, she will tell you which slice of cake the walnut is in. Mrs. Perspex is perfectly helpful. She can always guide you to the kernel of truth.

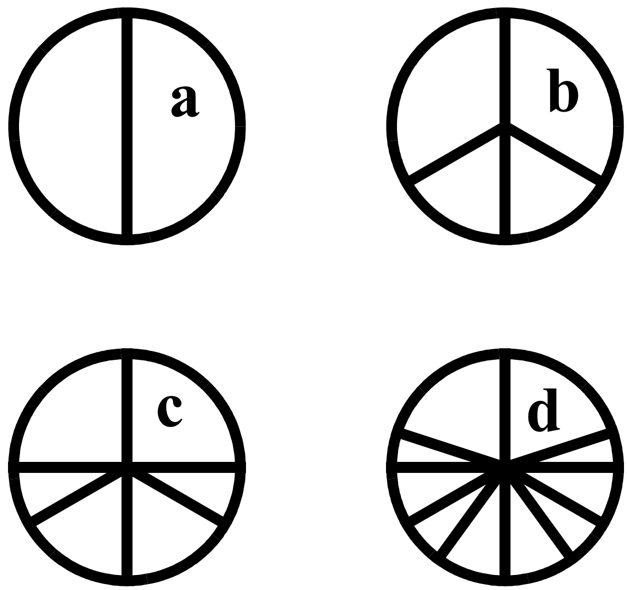


FIGURE 4. **Successive Rational Bounds (The Walnut Cake)**

In comes Mr. Turing, he is an apprentice baker and asks permission of you both to help by indicating the position of the walnut as well as he is able. Unfortunately, Mr. Turing is not perfectly helpful and is sometimes rather hesitant. Sometimes he is no help at all, but you have agreed to rely on his help. How are you going to cut the cake?

After a moment's reflection you decide not to cut the cake. You ask Mr. Turing if the kernel of truth is in the cake and he says that it is. You are just about to start eating the whole cake when in comes a friend of yours and asks for a fair share of the cake.

You decide to cut the cake in half, shown in diagram *a* in Figure 4 . You start from the centre and cut to the top of the cake. This defines the position zero on the circumference of the cake. If the kernel of truth lies at zero then Mr. Turing tells you that you have cut into it, otherwise he indicates that the kernel is somewhere in the rest of the cake. You then cut from the centre down. This defines the position $1/2$ on the circumference of the cake, clockwise from zero. The top of the cake is at a distance one, clockwise from zero. Again, Mr. Turing will tell you if you have cut into the kernel. Mr. Turing can tell if the kernel is at a computable position. You have cut the cake at zero, one half, and one. All of these are rational numbers and all rational numbers are computable. So long as you always cut the cake at rational positions Mr. Turing will always be able to tell you if you have cut into the kernel. If you miss the kernel then Mr. Turing tells you which of the two slices contains the kernel unless it is at an incomputable position in which case he hesitates for ever and never tells you anything. Earlier he was able to tell you that the kernel was in the cake because the circumference, from zero to one, is a segment of the number line, and Mr. Turing has arranged to stretch this segment of the number line onto the whole of the number line, so he knows that the kernel is somewhere along the line, even if it is at an incomputable position. But with two slices of cake Mr. Turing can never decide which slice contains the incomputable position of the kernel of truth.

In comes a second friend and asks for a fair share of the cake. You cut the cake into thirds as in diagram *b*. Mr. Turing can tell you if you have cut into the kernel, but will tell you nothing if the kernel is at an incomputable position. However, if the position is semi-incomputable and bounded above he will indicate all of the cake from a cut anticlockwise to zero. Conversely, if the position is semi-computable and bounded below he will indicate all of the cake from a cut clockwise to one. If the position is entirely computable then he will say which slice of cake the kernel lies in. He continues in this way as an infinitude of friends come in, one after another.

The largest slice of cake, that appears first, clockwise from zero, is shown with the letters *a, b, c, d,* and so on. If n people share the cake the largest slice has size $1/n$. If there are at least two slices of cake, then the smallest slice of cake is always adjacent to the largest slice. It has size $1/(n(n-1))$ which tends to the size $1/n^2$ as infinitely many people share the cake. If the kernel lies at a rational position you will eventually cut into it. If it is at an incomputable position then Mr. Turing will never tell you if you have found it. If it is at a semi-computable position you will usually end up with many slices of cake, one of which contains the kernel, but Mr. Turing will not tell you which one. If it is at a computable, irrational, position then you will have just one slice of cake which you can make as small as you like so that it contains the kernel and the minimum amount of sponge and walnut fondant. You can only be happy if the kernel lies at a Turing-computable position, whether it is rational or irrational.

Of course, what applies to a walnut cakes applies to anything you can measure with a ruler or compute with a Turing machine. If you go on refining the estimates in rational steps, which is the best refinement a Turing machine can accomplish, then measuring to a precision of $1/n$ will give an accuracy of at least $1/n$, but in some cases it will give an accuracy close to $1/n^2$. In general, the accuracy will be far higher than the precision. As you go on cutting the cake the number of different sizes of slice grows rapidly. When one hundred people share the cake there are more than one thousand different sizes of slice; so you can see there is quite a variation in the accuracy of measurements taken successively up to any given precision. If this surprises you then prepare for a shock.

Suppose that the kernel is close to $1/3$, that is it is close to the cut clockwise of the letter *b* in Figure 4 on page 54. If you increase the precision to quarters this gives you the less accurate estimates of $1/4$ and $1/2$. In general, increasing the precision of the best estimate so far makes the new estimate worse than the old one. For example, to be absolutely sure of confining the kernel within a thinner slice you must increase from n slices to n^2 slices. You have to work increasingly hard to be sure of improving the estimate. Almost every increase in precision gives an estimate that is worse than the best estimate so far. However, there are counter examples to this where only some of the rational numbers are used. If you limit the slices to a binary code, using just halves, quarters, eighths, and so on you can avoid the paradox of having the current cuts moving further away from the kernel. More bizarre ways to obtain the same effect are to arrange that, after the first denominator, all denominators are arbitrary, common multiples of their predecessors; or are arbitrarily chosen, but very much bigger then their immediate predecessor. All of these counter examples come at the price of using a less expressive number system than the rational numbers.

If we were only concerned with walnut cakes this would not matter, but the walnut cake theorem, and more sophisticated versions of it, applies to all scientific calculations, all engineering and physical measurements, and all searches for a word in a perspex brain. In almost every attempt to find a better estimate, measurement, or more accurate word, the attempt fails. This is one reason why scientific paradigms work for a while and then fail. With a little effort scientists can make good estimates and measurements and can express these in good theories written out in words. But almost every attempt to increase the accuracy of the estimates, measurements, or literary expressions, fails. The precision of a computation or measurement increases, and the number of words in a theory increases, but most of these increases in complexity are worse than the previous best theorem. Eventually, after at most n^2 work, a better paradigm is found and this becomes the new paradigm until, after n^4 work, a still better paradigm is found; and so it goes on, through n^8, n^{16}, n^{32} work, becoming ever more difficult to improve the current paradigm.

This might be frustrating, or exciting, for scientists, but it can be career limiting for politicians. The early politicians had it easy. Almost any quelling of the rabble populace was an improvement. Not so in Britain today. These days Government invites the populace to comment on proposed new laws, so as to try and scrape up ideas that might actually improve things. Conversely, Government sets up task forces, such as the Better Regulation Task Force, to remove excessive Government imposed regulations that make things worse, not better. These groups have the job of rolling back almost all of the regulations to get back to the few that worked well.

Ultimately there is a limit to the regulations that a human populace can bear, creating vast opportunities for cutting back on unproductive regulation or of moving regulation into computerised systems. It would be hard to evade tax if all monetary transactions were electronic and keyed to the DNA and fingerprints of the parties in a transaction. In this circumstance, we could repeal all tax laws and substitute a tax collecting program for them. We need not even have Government construct the tax program; it could be set free to evolve under the pressure of all those humans who want to receive public services. This might seem like science fiction, but one day human regulation of humans will grind to a halt under the growing burden of trying to make things better – and failing.

Scientists have examples of rolling back paradigms. Newton's theory of motion (1687) is pretty good, Einstein's theory (1915) is much better and far more complex, but in most practical cases Newton's theory is good enough. Engineers use Newton's laws, not Einstein's. They role back even the best available paradigm to one that is worse, but good enough.

The bureaucrats who try to regulate my scientific work have a hard time. The more they try to regulate, the more likely they are to waste effort and money. It is a reckless disregard of the walnut cake theorem to increase regulation without checking that unregulated, or less heavily regulated, institutions perform better. Let me say that again, for the benefit of judges engaged in a judicial review: *it is a reckless disregard of the walnut cake theorem to increase regulation without checking that unregulated, or less heavily regulated, institutions perform better.*

This is an example of Ockham's Razor. William of Ockham proposed that we should not increase the number of terms in a theory unless they make the theory better. For example, we should not increase the regulation of scientists, unless the regulation makes the scientists work better. In a universe like ours, where measurements and computations are limited to rational numbers, and thence to Turing-computable and semi-computable numbers, the utility of Ockham's Razor is proved by the walnut cake theorem.

Thus, we see that the walnut cake theorem explains what paradigms are, why paradigm shifts occur, and why Ockham's razor works. But on a human level it does much more than this. It explains why robots will share the human condition of making errors. A perspex machine, limited to rational numbers, will always make errors, even if it operates perfectly, as shown by the walnut cake theorem. Only a continuous, computing machine, such as the pin-hole, perspex, machine can operate smoothly, without ever requiring a paradigm shift. But in our universe, it seems unlikely that a perfect, continuous, machine can be built so every kind of machine will suffer error. Even a perfectly operating, continuous, perspex machine would have to be extraordinarily lucky, or contain a time machine, if it were to get every prediction, measurement, and expression right. No scientist or politician can operate without error because they do not have access to time machines, and because they communicate in words, which are discrete symbols, not continuously varying quantities. Even if they could communicate with continuously varying quantities the physics of the universe in the present moment would, presumably, cause them to be in error.

Whatever choice we make of the numbers a perspex machine uses it will either have less expressive power than a rational perspex machine, and of us, or else it will suffer error. Error is inevitable in any sufficiently expressive, symbolic language that describes the world, and whose users do not use a time machine.

Public Language

Let us return to our imaginary android who has been learning sensorimotor sentences by playing with a cup. Humans at birth are able to hear about 200 speech sounds, or phonemes. Let us suppose that the android is equipped with a way of hearing these phonemes by reading perspexes from a perspex, auditory cortex and of reproducing phonemes by writing perspexes into a perspex, vocal cortex. These cortexes contain a continuously varying manifold of sensed and produced sounds, with flat neighbourhoods, like words, marking out the prototypical sounds of particular speakers, including the android itself.

The parent puts a cup in the android's hands and says, "cup." The phonemes in "cup" grow a sensory model in the auditory cortex in exactly the same way as the positions of the android's fingers grow a sensory model in the motor cortex. If we suppose that the android grows a bi-directional relationship between the motor and auditory cortexes, then it can pass from hearing "cup" to sensorimotor sentences describing the cup. If it writes perspexes into its vocal cortex it can produce phonemes and relate these to the other models of the cup. Of course, we would want to prevent the android from relating every model in its mind to every other model. This might be done by reinforcing connections according to the amount of time they are used. The perceptual connections are used only in the presence of a physical stimulus so this goes some way to providing the required reinforcement of significant perceptions, and, conversely, the neural pruning of unused connections, which fade away to some base level, which we might set to zero.

In this scheme it is important for the parent to play with the android, presenting it with socially significant objects, words, and actions. The android too needs time to play, in order to calibrate its auditory and vocal cortexes by talking to itself and checking that these sentences refer to its internal, sensorimotor sentences. We may well suppose that the android will find that it can process sentences faster, and to its social benefit, by manipulating the public language sentences internally, and refraining from speaking them. Thus, it is driven by a causal chain of deterministic physical and social pressures to develop an internal language that is tuned to the needs of communication with other androids and humans.

The android, perhaps being the first of its kind, or an isolated example of its kind, or a creative example can produce phonemes in the absence of a parent. If it writes any perspexes, such as the first few in a sensorimotor sentence, into its vocal cortex it will produce sounds that are causally related to the sensorimotor sentence. Thus, the physics of the universe imposes a deterministic link between an internal sentence and a spoken one. It is then a small, but very significant, step to imagine that the evolutionary pressures on an android, and any other androids or humans, will cause everyone to enter a linguistic arms race, shaping the length, similarity, grammar, and frequency of spoken words, so that words relate to both android and human internal, sensorimotor languages. If we are to speak to androids and fully understand them, then their bodies and minds must be similar to ours.

Let us look at a little of the English language to see how it might relate to the kinds of visual concepts a perspex android might have. This might give us some idea of the physical properties, such as eye colour, that we should build into an android. The appendix *Visual Phrases* contains a short list of English phrases that are related in some way to the human sense of vision.

Some of the phrases refer to physical objects: *blue eyed, brown eyed, eyeball, eyeless, grey eyed, one eyed, skew eyed, wall eyed, within eyesight, in plain sight.* An android which is nominally possessed of two, well calibrated, eyes that have coloured irises, might be able to relate sensorimotor meanings to these words that are practically identical to the meanings held by a human.

Some of the phrases refer to physical behaviours associated with seeing: *clap eyes on, eyeball, get one's eye in, in the blink of an eye, in the wink of an eye, meet the eye, out of the corner of the eye, poke in the eye, rivet one's eyes, exchange glances, glance, sideways glance, quick glance, look askance, look over, look through, on looker, peek, peep, pore over, review, rubber necking, scan, sight-see, catch sight.* An android that has social experience of itself, other androids, and humans might be able to see that androids and humans share these behaviours. It might, therefore, be able to arrive at a similar meaning for these phrases to the meaning held by humans.

Some of the phrases refer to physical limitations or physical side effects: *bleary eyed, colour blind, snow blind, blind, blind as a bat, blind side, blind spot, blinding, blinding light, blindfolded, black eye, eagle eye, naked eye, pink eye, red eye, sharp eye, unaided eye, eye strain, make out, unseeing, long sight, near sight, short sight, weak sight, sight, sightless.* An android might be able to relate the human meaning of these phrases to itself to the extent that its body is subject to the same limitations and frailties as the human body. Alternatively, it might understand the human meaning by observing the behaviour of humans and listening to their reports of limitation and suffering.

Some of the phrases relate to feelings, emotions, or psychological motivations: *blind hope, blind faith, evil eye, eye-catching, eyeful, keep half an eye on, glint in the eye, poke in the eye, gape, gawk, gawp, furtive glance, inviting glance, sideways glance, glare, goggle, leer, dirty look, ogle, peek, peep, beautiful prospect, exciting prospect, fine prospect, good prospect, high prospect, low prospect, sight see, scowl.* We consider how a perspex android might have appropriate motivations in the chapter *Intelligence*, and how it might have appropriate feelings and emotions in the chapter *Feeling*.

Some of the phrases relate to analogical views within the speaker's mind's eye: *blindingly obvious, blindingly stupid, blindfolded, blinkered, hit between the eyes, keep half an eye on, use your eyes, look at it my way, look at it from my point of view, look at it from my standpoint, look out for oneself, look to the future, look to the past, look to your family, look to your honour, look to your sword, from my perspective, keep a perspective, oblique perspective, beautiful prospect, exciting prospect, fine prospect, good prospect, high prospect, low prospect, regard, high regard, low regard, no regard, review, revision, see here, see beyond the end of your nose, see into the seeds of time, see it my way, see it through, see it to the end, see my way clear, see red, see through a glass darkly, see through rose tinted glasses, see through the veil of tears, see to the heart of the matter.* A perspex android has the potential to turn any computer program into a 3D geometrical model, so it can have an actual view of all of the above things. For an android "blindingly stupid" might simply refer to the fact that the android is being stupid because it is blinded by a perspex that prevents it from seeing an actual, clear way forward to its goal, when a small shift of perspective would bring its goal into clear view.

No doubt there is a great deal more that can and should be said about how a perspex robot might acquire language and use it in a mixed society of humans and robots, but let us return to the central thesis of this chapter. Visualisation is more than language.

Beyond Language

In this chapter I have said very little about the grammatical structure of language or the phonemic structure of words. I have, instead, concentrated on how a perspex android might acquire internal, sensorimotor languages, and how it might relate these to a public language following a causal chain determined by the physical constraints of the universe, including the constraints of the android's own body and brain, and the constraints of living in a mixed society of humans and androids.

I crated the walnut-cake theorem to show how language leads to the need to change paradigms. The theorem explains why Ockham's razor is a good guide to conducting science. I have touched on the political regulation of society and how androids, or at least computers, might play a role in improving society. In short, I have alluded to the interesting things an android might say, if it possessed language. In the absence of language we must communicate with an android, or with an extraterrestrial being, the same way we do with animals – by very careful ethological study and experiment.

At one level, language is defined by the manipulation of symbols so it is interesting to consider what properties are manifest in a symbol. The Gödel proofs assure us that a Turing machine could use nothing but the written form of integers as its symbols, and claims that whatever symbols it does use can be translated into integers. This is true, but the translation has to be capable of being performed by a Turing machine. A Turing machine is committed to being able to deal with Turing-computable numbers by the symbols it uses. To see this, consider the counter case where all integers are related to Turing-incomputable numbers by a one-to-one correspondence, or isomorphism. The Turing-incomputable numbers can, theoretically, be translated into their corresponding integers, but there is no way for a Turing machine, or a human mathematician, to do this. The theoretical existence of an isomorphism is not enough.

A Turing machine is defined to read a symbol and distinguish it from all other symbols in one unit of time. The symbols have to be discrete, as Turing defined. The task of distinguishing symbols implies all of the results of Turing computability. Turing computability is a manifest property of Turing symbols. Conversely, if a machine could read a continuous symbol and distinguish it from other continuous symbols then it would be able to read and distinguish Turing-incomputable numbers. This is how I defined the perspex. It can read matrices of real numbers, and can compare the magnitude of any two real numbers, even if these are Turing-incomputable. In theory, the ability to do more than any symbolic system, including mathematical logic and language, is a manifest property of the perspex.

It is conceivable that an extraterrestrial being might communicate in the continuum and might not use any language at all. This would make finding radio signals from extraterrestrial, intelligent, life forms even more difficult than is currently supposed. However, such extraterrestrial beings might be capable of restricting themselves to a linguistic form so that they can communicate with us. Alternatively, they might communicate via a continuous, meaningful form, such as sign "language." Strictly, sign languages are not just symbolic languages because the geometrical extent of a sign, co-articulation of sign actions, and speed of signing are meaningful. In the same way, spoken language is not entirely symbolic. The continuous aspects of tone of voice, co-articulation, and speed, carry speech into the continuum. Humans do not behave as if they are entirely symbolic systems, so there is some hope that we could communicate with an extraterrestrial being that does not use any kind of language.

In practical terms, we need not go so far as to construct a perfect, continuous perspex machine in order to explore the implications of the geometrical arrangement of words and actions laid out on a manifold. We could construct a Turing-computable approximation to these things using a conventional, digital computer. In essence this would be a linguistic task, but it would simulate a visual one.

Conversely, a consideration of how a visual system might simulate a linguistic definition of consciousness leads us to a practical definition of visual consciousness. This practical definition allows us to propose how a robot might be constructed so that it has free will, intelligence, feeling, a perception of time, and an engagement with spirituality. These are all things we want an android to experience and talk to us about.

Questions

1. The Unlimited Register Machine is equivalent to a Turing machine, but is very like a standard computer language. Students find it easy to understand and use, so they make far more progress on practical and theoretical problems than by using the Turing machine. If you are a lecturer, why not teach computability mainly in terms of the URM? If you are a student why not translate your lecturers notes on the Turing machine into the URM? If you are one of my students, tough guys, you will use the perspex machine – but you are free to translate it into anything you like!

2. If you are a computer science student you may well have noticed that the 2D perspex machine implements a two-address code and the 3D perspex machine implements a three-address code. Can you make a better job than I have done of implementing these codes in a perspex machine? What does a compiler optimiser look like when it is implemented in perspexes? Which parts of an optimiser look like biological DNA enzymes? Which redundant parts of perspex programs look like prion-induced tangles, or tumours in an animal brain?

3. If you are a French speaking, mathematics student you may well have noticed that the 2D perspex machine is the more likely candidate for axiomatisation[1,2]. Can you arrange that the axioms of projective geometry are a subset of your axiomatisation of the perspex machine? Can you find any interesting dualities between perspex programs and objects in projective geometry? If so, what are the program duals of the classical theorems of projective geometry? If you succeed in this, can you arrange that the perspex machine performs the whole of human mathematics, and more? If you succeed in this, how should the Bourbaki treat your axiomatisation? By then, you will probably be a member of the Bourbaki.

4. Which results of number theory bear on the issue of paradigm shifts?

5. What does the super-Turing status of the perspex machine say about logical necessity?

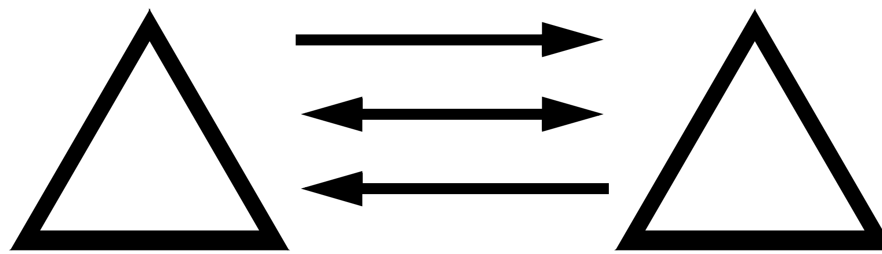
Further Reading

1. **Anderson, J.A.D.W.** “Representing Visual Knowledge” in *Philosophical Transactions of the Royal Society* vol. 352, pp. 1129-1139, (1997).

2. **Anderson, J.A.D.W.** "Perspex Machine" in *Vision Geometry X1*, Longin Jan Lateki, David M. Mount, Angela Y. Wu, Editors, *Proceedings of SPIE* Vol. 4794, pp. 10-21, (2002).
3. **Anderson, J.A.D.W.** "Exact Numerical Computation of the General Linear Transformations" in *Vision Geometry X1*, Longin Jan Lateki, David M. Mount, Angela Y. Wu, Editors, *Proceedings of SPIE* Vol. 4794, pp. 22-28, (2002).
4. **Boolos, G.S. & Jeffrey, R.C.** *Computability and Logic* Cambridge University Press, (1996). First Published 1974.
5. **Chomsky, N.** *Syntactic Structures* Mouton, (1957). Reprinted 1985.
6. **Cotogno, P.** "Hypercomputation and the Physical Church-Turing Thesis" *British Journal for the Philosophy of Science,* number 54, pp. 181-223, (2003).
7. **Coxeter, H.S.M.** *Projective Geometry* 2nd edn., Springer-Verlag, (1994).
8. **Cutland, N.J.** *Computability: an introduction to recursive function theory* Cambridge University Press, (1980).
9. **Dreyfus, H.L.** *What Computers Still Can't Do: a critique of artificial reason* MIT Press, (1997). First published 1972.
10. **Nagel, E. & Newman, J.R.** *Gödel's Proof* Routledge, (1993). First published 1958.
11. **Popper, K.R. & Eccles, J.C.** *The Self and its Brain - an argument for Interactionism* Routledge, (1995). First published 1977.
12. **Pour-El, M.B. & Richards, I.J.** *Computability in Analysis and Physics* Springer Verlag, (1989).
13. **Swartz, R.J.,** editor, *Perceiving, Sensing, and Knowing* Anchor Books, (1965).
14. **Turing, A.M.** *On Computable Numbers, With an Application to the Entscheidungsproblem* Proceedings of the London Mathematical Society, series 2, volume 42, part 3, pages 230 – 265, (1936).
15. **Turing, A.M.** *On Computable Numbers, With an Application to the Entscheidungsproblem. A Correction.* Proceedings of the London Mathematical Society, series 2, volume 43, part 7, pages 544 – 546, (1937).
16. **Turing, A.M.** *Computing Machinery and Intelligence* Mind, volume LIX, number 236, pages 433–460, (1950).
17. **Weihrauch, K. and Zhong, N.** "Is Wave Propagation Computable or can Wave Computers Beat the Turing Machine?" *Proceedings of the London Mathematical Society,* number 85, part 3, pp. 312-332, (2002).

CHAPTER 7

Visual Consciousness



Introduction

On the one hand consciousness is supposed to be the most difficult thing to understand. Some philosophers suppose that there is something ineffable in the feeling of redness that accompanies the sight of something red. Some philosophers suppose that there is an uncrossable explanatory gap between the physical things that go together to make a brain, and the non-physical, subjective, experiences in a mind. Some philosophers suppose that mind is non-physical, or non-spatial, so that there is no way at all for science to understand mind. On the other hand, we can look up "consciousness" in a dictionary[6] and read: "aware of one's surroundings, one's own thoughts and motivations, etc." I will readily admit that I do not know how to program a general "awareness" into a computer, but I do know how to program a computer so that it can see. If we replace "awareness" by "seeing" in the above definition, and add a few more terms, we arrive at some of the properties of visual consciousness[3]. Visual consciousness is seeing one's surroundings, one's own thoughts, motivations, feelings, and so on. This definition allows us to add new properties to be visually conscious of, and does not exclude other kinds of consciousness. We might program consciousness in auditory, olfactory, haptic, proprioceptive, kineceptive, and all manner of sensory modalities, each in their own kind. Alternatively, we could display each of these sensory modalities in pictorial form, and thereby use visual consciousness to be conscious of each of these sensory modalities. Thus a robot might hear trumpets as bright red, or taste sugar as yellow, or feel sandpaper as sparkling. This sensory jumble, or *synaesthesia*, is regarded as a mental illness when it occurs in humans, but it might be a valuable step toward integrating all manner of sensory modalities of consciousness in a robot. Robots might well start off as mentally handicapped members of society before we develop their mental abilities to match the wide range of our own.

Before we embark on this wide ranging enterprise, I choose to identify the most elementary property of visual consciousness so that I can try to manifest it in the myriad ways desired. I claim that to be visually conscious is to have a partial, bi-directional relationship between perspexes. If this is accepted, it gives a very simple resolution to the problems of ineffability and the explanatory gap. Ineffable experiences are those that are not Turing-computable, and the physical content of an experience closes the explanatory gap.

In the chapter *Free Will* we see how it is causally necessary that a robot develops individuality, and in the chapter *Intelligence* we see how it might develop consciousness of self. In the chapter *Feeling* we see how the explanatory gap is closed in the experience of the timeness of time, before exploring the redness of red, and the physical nature of other sensory qualities. Here we concentrate on the nature of visual consciousness and its relationship with other kinds of consciousness.

Visual Consciousness

For millennia philosophers have debated the precise definition of the concepts of *knowledge*, *seeing, and consciousness*. Around 350 BC Aristotle said, in effect, *to see is to know what* is *where in an image*. Let us now explore the idea of visual knowledge[2] and see how it becomes visual consciousness[3]. Once we have identified the essence of visual consciousness we will be in a position to say what the essence of any kind of consciousness is.

Firstly, let us examine the concept of an *image*. We all know what images are. Images form on our retinas, appear in mirrors, cameras, photographs, televisions, computers, paintings, drawings, tattoos, printed things, embossed things, carvings, holographs, chromatographs, sonographs, siesmographs, embroidery, mosaics, formal gardens, plant and animal markings and, no doubt, many more things that do not spring immediately to mind. The common element to all of these is that there is a spatial arrangement of some quantity. For example, the density of black ink in a newspaper photograph varies across the page, and the intensities of red, green, and blue light vary across a television screen. The former of these is what physicists call a *scalar field* and, at moderate spatial resolutions, the latter is a *vector field*. I would allow that any field, that is, any spatial arrangement of any kind of quantity, is an image. I am happy to agree that a human may see an image formed by light falling on the retina of the eye, and that we may equally well see an image formed by pin pricks arranged in an array on our backs. I draw no philosophical distinction between the different physical processes that are involved in forming an image. Specifically, I will allow that the spatial distribution of electric charge in a camera, or in computer memory forms an image which can be seen by a robot. For me, *any field is an image* in the technical sense of the word "field" as used by physicists.

Secondly, let us examine the concept of *knowing what is in an image*. Figure 5 shows a drawing of a small section of a flower bed in a garden. Imagine being asked, "What is that?" when the questioner indicates the part of the garden shown here at the centre of the black circle.

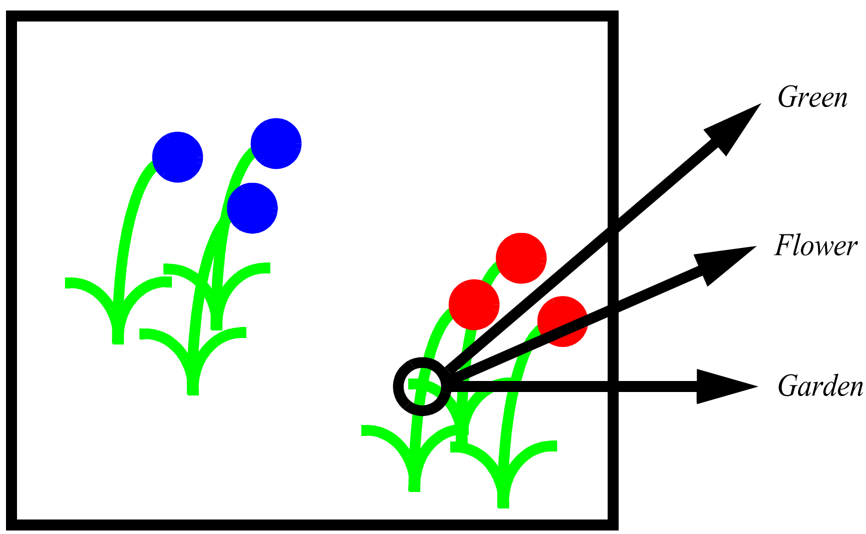


FIGURE 5. Knowing What is in an Image

If you had recently been discussing painting you might answer, "It is green." Alternatively, if you had been discussing plants you might say, "It is a flower." Or, if you had been discussing a house you might say, "It is the garden." Indeed you might have all of the concepts *green*, *flower*, and *garden* in mind when you answer in any of these ways. That is, all of these concepts might come to mind when that part of the image is indicated, but you might choose to speak about any, all, or none of them. What you say does not affect your knowledge of what is in an image, though a scientist would want some verbal, or other kind of, report from you as evidence that you do know what is in the image.

The arrows in the figure indicate a *directed mapping,* or *directed relationship,* that runs *from* the part of the image in discussion *to* each of the concepts. The existence of even one relationship in your mind would demonstrate that you know something about what is in the image. The existence of many relationships would demonstrate that you know many things about what is in the image of the garden. Without digressing into the theory of knowledge expounded by philosophers let us admit that humans are often wrong in what they claim to know, and can know things other than facts. As an example of non-factual knowledge, consider the case where you cut the flower at the point indicated and say, “It’s for you,” and smile sweetly. The relationship in your mind might then be from the image to knowledge of how to cut flowers and give gifts. As discussed in the chapter *Beyond Language*, the physical acts of cutting and giving might involve continuous actions that are Turing-incomputable. Such actions cannot be put exactly into words, or exactly into any kind of symbolic fact.

For our purpose of designing robots with visual knowledge and visual consciousness we may lump all kinds of knowledge, belief, assumption, and skill, together in a perspex program. For example, we may express the knowledge that the *green* thing is a *flower* or a *garden* in a perspex program that allows a robot to work with these concepts. Equally, a perspex program can perform the physical actions of cutting flowers, giving, and smiling, even if these actions are Turing-incomputable. Thus a perspex program will suffice as a representation of all kinds of knowledge for the purpose of defining visual knowledge.

Thirdly, let us examine the concept of *knowing where something is in an image*. Figure 6, on page 69, shows where the concept *green* is portrayed in Figure 5, on page 67,. Green occurs at the precise point under discussion and in all the other points coloured green in Figure 6. Thus, we may imagine further arrows coming from the concept *green* to every green pixel in Figure 6. However, this is a slightly mistaken impression of where green occurs in Figure 5 because one of the flowers is missing a leaf.

Now consider where the flower is, Figure 6 shows the location of the green stem of the flower under discussion. We would not accept any other flower as proving that a human observer knows where the flower is because we expect humans to be able to indicate positions with sufficient accuracy to show which of the six flowers is under discussion. However, we know that humans are not infinitely accurate and might miss the considerable displacement that we could measure by placing a tracing of Figure 6 on top of Figure 5. The whole of Figure 6 is displaced upward and to the left, with respect to the black box forming its frame. If you were not deceived by this, can you tell how Figure 7, on page 70, is laid out?

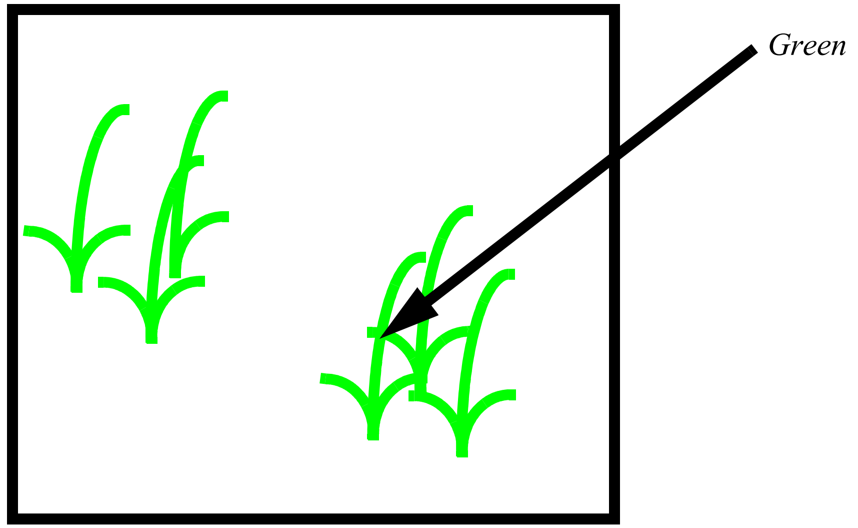


FIGURE 6. Knowing Where *Green* is in an Image

Now consider where the garden is. The observer might indicate the whole of the garden shown in Figure 5. We would allow this as proving that the observer knows where the garden is, even though we know that the figure shows only part of the garden (as stated in the introduction to these examples).

By way of further illustration of this point, imagine standing in a garden and looking beyond your feet. What do you see? If you say, “The Earth,” I will agree with you, even though you can see only a tiny part of the Earth. No matter how tall you are, even to the extent of having your head in outer space, you cannot see the whole of the planet Earth. It is a brute mathematical fact that there is no where in three dimensional space where you can see the whole of the planet Earth at the same time. But seeing just part of an object, and identifying it is held to be seeing the object.

Thus we see that, in human terms, the precise geometry of the thing indicated is not essential to knowing where it is. We know that we operate with limited accuracy, with limited memory, and limited perception. The defining characteristic of knowing where something is is to have a directed relationship in mind from the concept to some part, all, or more than the extent of the thing in an image. The *where* part of Aristotle’s definition of *seeing* requires that we can indicate a location in an image.

Drawing all of this together, we can see that knowing *what* something is involves a directed relationship from an image to a concept, but it need not be spatially accurate. Knowing *where* something is involves a directed relationship from a concept to an image, but it, also, need not be spatially accurate. Given these mismatching inaccuracies in both directions, we say that visual knowledge is knowledge that is in a partial, bi-directional relationship with an image. Here it should be understood that being a partial relationship does not preclude the possibility of being a total relationship, in which case the visual consciousness is spatially exact.

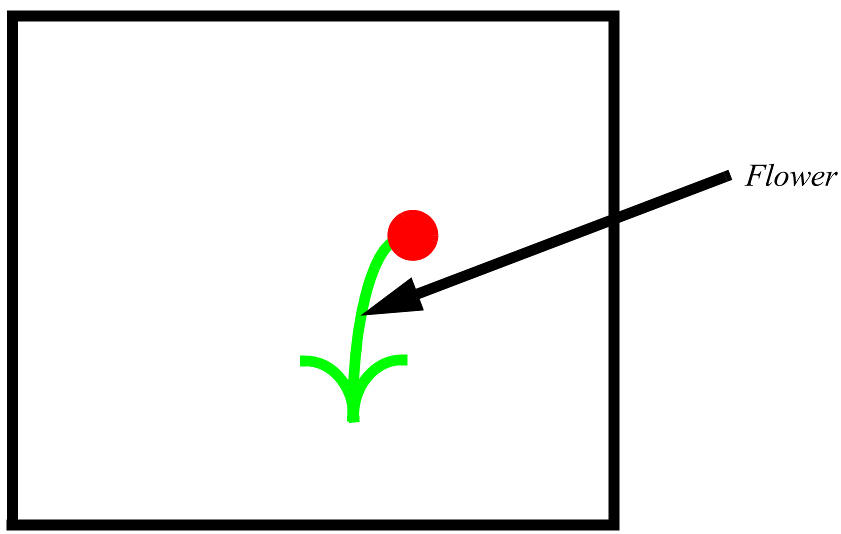


FIGURE 7. Knowing Where a Flower is in an Image

However, we are dealing with perspexes so knowledge represented as a perspex program can be seen as a geometrical simplex in an image, and physical things can be regarded as perspexes. Hence visual knowledge is, in essence, just a partial, bi-directional relationship between perspexes.

Further, it seems to me that if a human being can say what is in an image, and can indicate where a thing is in an image, then he or she is conscious of the thing. Conversely, if the human being can say what is in an image, but cannot say where it is then I would diagnose the mental deficit of *blind sight*[10], and say that he or she is not conscious of the thing. Similarly, if a human being can indicate where a thing is in the image, but cannot say, or otherwise indicate, what the thing is I would diagnose the mental deficit of visual agnosia[10], and say that he or she is not conscious of the thing. I claim that the test of visual consciousness is to demonstrate visual knowledge. If visual consciousness has any further properties beyond the possession of visual knowledge, then we might discover what these are in subsequent discussion, but, for now, I propose the definition that visual consciousness is visual knowledge. In other words, visual consciousness is a partial, bi-directional relationship between perspexes.

All Possible Uses of Consciousness

Some philosophers wonder what use consciousness is. What is it for? What is to stop a being operating according to unconscious measurements of the world? If you accept the definition just given, that visual consciousness is a partial, bi-directional relationship between perspexes, then there is no mystery in visual consciousness, and hence, by synaesthesia, no mystery in any modality of consciousness. There is also no mystery as to what consciousness is for. In the absence of partial, bi-directional relationships a program could scarcely relate to the world at all.

Imagine that there is a hungry tiger hiding in the long grass, and that you are a fat human being. You will make very good eating, if the tiger can catch you. In order to survive it is vitally important that you know *what* is in the long grass, and *where* it is. Nature has ensured that you are the fastest mammal over a distance of six feet. Being the fastest mammal on Earth over one, human, body length is not enough to outrun a tiger, but if you had the presence of mind to bring a sharp stick with you, it is enough to side step the tiger, poke it in the eye, and dine on tiger steaks. Knowing *what* is *where* is visual knowledge, and hence visual consciousness. Visual consciousness keeps you alive. By contrast, if you do not know what is in the long grass there is little reason to prepare for an attack, even if you know precisely where the long grass is. Similarly, if you know there is a tiger nearby, but do not know where it is it will be difficult to react quickly enough to save yourself from the tiger's deadly pounce. The absence of any part of visual consciousness is life threatening under these circumstances. Quite simply, having visual consciousness gives you the best possible chance of staying alive; that is what visual consciousness is for.

The same is true of consciousness in any other sensory modality; knowing what is where in the world increases our ability to interact with the world and stay alive.

Some philosophers claim that they can be conscious in a non-spatial way: that they can know what something is without having any knowledge of where it is. For example, they suppose that they can know that the number $\sqrt{2}$ is irrational without knowing anything spatial. But what use can a philosopher make of this knowledge unless he or she knows *where* this abstract knowledge will be useful? Will it be useful in a particular mathematical proof, say at line 26? Will it be useful to impress a sexually attractive guest at a cocktail party? Will it apply here and now to this problem, or somewhere, some time to another problem? If non-spatial consciousness can exist I suggest it is of no use unless it can be applied somewhere in the physical universe, thereby obtaining a *where* part eventually.

If a philosopher knows that the number $\sqrt{2}$ is irrational then he or she has a relationship from the number $\sqrt{2}$ to *irrationality;* this provides the *what* part of knowing. But if this knowledge is to be non-spatial, having no *where* part, the number $\sqrt{2}$ must be non-spatial, and the philosopher must be unable to write the spatially arranged symbol "$\sqrt{2}$" when asked to give an example of an irrational number, and must not be able to point to any person or thing that says by voice, sign language, or any kind of physical symbol, "Root two is irrational." There might be such philosophers, but they cannot consistently write, "… there are no spatial concepts at all that apply to … $\sqrt{2}$" Clarke[5] page 167, quoted on my page 80, defeats his own argument in this way by writing spatially on the page.

But what difference does it make if non-spatial kinds of consciousness can exist? If we have visual consciousness, we can always ignore the *where* part of visual consciousness and supply the required kind of non-spatial, *what* consciousness; or we can ignore the *what* part of visual consciousness and supply only a *where* consciousness. Or we can ignore both the *what* and the *where* parts of consciousness, and provide a Zen-like experience of nothing related to nothing. Visual consciousness provides, at least, the relationships needed by any possible kind of consciousness. It can, therefore, support all possible uses of consciousness.

All Possible Conscious Feelings

Some philosophers argue that feelings are described completely by the relationships they bear to ideas. If this is so then, as we have just seen, visual consciousness can supply all of these functional kinds of consciousness.

Some philosophers argue that feelings have a physical content, distinct from functional relationships. If so, perspexes can supply, say, the physical passage of time, as well as the functional perception of time. Further examples of the physical contents of feelings are discussed in the chapter *Feeling*. The special nature of time is discussed in the chapter *Time*.

Whatever physical content might be wanted of a feeling can be provided by perspexes – following the hypothesis that perspexes can describe the physics of the entire universe by describing the geometrical structure and motion of everything. If this proves to be false then I reserve the right, as set out in the chapter *Perspex Instruction*, to re-define perspex machines in terms of the unaccounted-for physical phenomenon, so that a perspex machine can, eventually, match the computational power of the universe.

Some philosophers argue that feelings have a non-physical content. This is denied by materialism. My body is part of the physical universe, according to the philosophical precept of materialism. But my body is physically related to every other part of the universe. If there is something that is not physically related to my body it is not part of the universe that contains my body, but is part of some other universe. Therefore, if non-physical feelings exist they do not affect my body in any way; they are entirely undetectable and utterly useless in this universe. I deny that any non-physical thing exists, but if it can be proved to me that they do then the proof relates that thing to the physical universe and it is then, by definition, a physical thing. No philosopher can consistently argue, using any physical means, for example, speech or writing, that there is any non-physical thing. This theme is taken up, amongst others, in the chapter *Spirituality*.

Thus, I argue, the perspex machine can, in principle, experience all possible conscious feelings.

Gradations of Consciousness

We have already seen gradations of functional consciousness by knocking out the *what* and/or *where* parts of visual consciousness. We can imagine having more or less consciousness depending on the number of partial bi-directional relationships that are maintained in a mind.

We have discussed gradations of consciousness by having some range of physical contents. We can imagine having more or less consciousness by having different numbers of physical contents to feelings.

We can imagine having different intensities of conscious feeling by having more functional relationships, or more physically intense contents. The passage of a minute of time can be a more intense experience than the passage of a second of time, because it has more physical content.

We can imagine gradations of feelings by virtue of being attended to, more or less strongly, by the self. The concept of self is taken up in the next section and in the chapters *Free Will* and *Intelligence*.

There is a practical, scientific, reason to want gradations of consciousness. We can start by providing a robot with the simpler, weaker sort, whichever they are, and add to the gamut and intensity of a robot's conscious experiences as the know-how develops. If consciousness is an all-or-nothing kind of thing we might never understand it, and might never be able to find out how to implement it in a robot. This is one reason why philosophy is important to science. Philosophy provides a range of possible kinds of consciousness which the scientist can pick off and try to implement in a robot.

Consciousness of Self

If a being is to be conscious of self, there must be a self to be conscious of.

Individuality arises automatically in a being that has free will, simply by the accretion of the consequences of the choices made, as described in the chapter *Free Will*, but this does not automatically lead to a concept of self.

The concept of a singular being arises from the physical necessity of having one's body do one gross thing at a time. We can side step to the left and stab the tiger, or side step to the right and stab the tiger, but our body cannot do both. If we are to interact successfully with the world some concept of the singularity of body must be embedded in a being's mind. A plant can grow to the left, or grow to the right, or follow the sun, but a single stem cannot grow in more than one volume of space at a time. Even plants have a concept of singularity built into them. But a concept of singularity of body does not lead automatically to a concept of self.

If a being is sufficiently intelligent, there is survival value in being able to picture its body in different physical arrangements: side stepping to the left; side stepping to the right; or considering that we might meet a tiger today, sharpening a stick, and taking it to help us push through the long grass. Intelligence does lead to a concept of self, as I will argue in the chapter *Intelligence*.

A being can be conscious of self to different degrees, because there are myriad physical contents and functional ways to be conscious. There are many physical contents to a body and its interactions with the universe, many physical intensities in a body and its interactions, and many functional relationships in a highly intelligent being.

Once a being is possessed of a concept of self it can define itself in relation to others in a society, or in relationship to a deity. These issues are taken up in the chapter *Spirituality*.

Against Accounts of Non-Spatial Minds

Having set out a positive, philosophically and scientifically testable, hypothesis of what constitutes visual consciousness, it remains to argue against those who maintain that this enterprise is hopeless. In doing this, as in any philosophical argument, the task is to enter into the opposing argument, as far as possible, to see what it reveals about our own position. The most extreme antithesis to our position is to deny that mind has any spatial properties. Here I examine two proponents of this position: McGinn[7] who accepts that brains exist in spacetime, but argues that mind exists only in time and not in space; and Clarke[5] who argues that mind existed before spacetime and now causes spacetime as the observer of a quantum universe. No doubt there are other antitheses that should be considered, but for the purposes of introducing a rebuttal defending the hypothesis of visual consciousness these two will suffice.

McGinn: Consciousness and Space

In criticising McGinn[7] I quote what I take to be the strongest paragraphs of his text, in the order they occur, and analyse each in turn. I then synthesis my individual responses and assess the impact of this dialectic both on McGinn's thesis and my own. The quotations are set out with the original spelling, punctuation (except that hyphens might be introduced to justify the text) and font families, such as italic. Thus I remain true to the original text in all important particulars, notwithstanding the fact that my choice of excerpts is open to challenge.

> "It is hard to deny that Descartes was tapping into our ordinary understanding of the nature of mental phenomena when he formulated the distinction between mind and body in this way – our consciousness does indeed present itself as non-spatial in character. Consider a visual experience, E, as of a yellow flash. Associated with E in the cortex is a complex of neural structures and events, N, which does admit of spatial description. N occurs, say, an inch from the back of the head; it extends over some specific area of the cortex; it has some kind of configuration or contour; it is composed of spatial parts that aggregate into a structured whole; it exists in three spatial dimensions; it excludes other neural complexes from its spatial location. N is a regular denizen of space, as much as any other physical entity. But E seems not to have any of these spatial characteristics: it is not located at any specific place; it takes up no particular volume of space; it has no shape; it is not made up of spatially

> distributed parts; it has no spatial dimensionality; it is not solid. Even to ask for its spatial properties is to commit some sort of category mistake, analogous to asking for the spatial properties of numbers. E seems not to be the *kind of thing* that falls under spatial predicates. It falls under temporal predicates and it can obviously be described in other ways – by specifying its owner, its intentional content, its phenomenal character – but it resists being cast as a regular inhabitant of the space we see around us and within which the material world has its existence. Spatial occupancy is not (at least on the face of it) the mind's preferred mode of being."

Consciousness and Space page 97.

McGinn starts with the assertion that Descartes was describing the normal, human, condition when he said that, to the owner, body feels spatial, but mind feels non-spatial. This might be true of the majority of humans, but I am not aware that I have any non-spatial thoughts. However, I do seem to remember that before written words were integrated in my mind, in my mid-thirties, the meanings of written words were so ephemeral as to defy being pinned down in space, or being related to anything very much. Perhaps I am mistaken in this, which would not be surprising, because introspection is such an unreliable method of psychological analysis. For the present, I simply note McGinn's claim that humans have non-spatial, conscious thoughts.

McGinn then gives an example of what he regards as a non-spatial thought in the visual experience of a yellow flash. He acknowledges that the experience correlates with changes in the spatial structure of the brain, but claims that the experience of the yellow flash is non-spatial in several particulars. Firstly, that it is not located in space. I do not know how McGinn experiences flashes, but for me they occur either at an isolated point or region in my field of view, or else they cover the whole of my field of view. I can also create yellow flashes in my mind's eye in which case they can occur at an isolated point, or in a region of a space of some modest dimensionality. In either case I experience flashes in a spatio-temporal way. They do fill out a particular volume of space and have a shape. Large flashes are made up of the spatially distributed parts that, in other circumstances, would be filled by smaller flashes. Flashes in the world are two dimensional when they fill my field of view, and three dimensional when I can see the source of the flash, such as the flashing light on a police car. Furthermore, flashes in my mind's eye are, typically, no more than four or five dimensional, though I can maintain a very simple seven or eight dimensional spatial configuration. Flashes are solid if they fill out a space of three or more dimensions. Frankly, I cannot conceive how McGinn consciously experiences flashes if it is not like this, however, I am prepared to take him at his word.

McGinn then gives an example of a non-spatial thing. He says that numbers do not have spatial properties. This is wrong. Octernions fill out an eight dimensional space, but I am not aware that they have any practical application. Quaternions fill out a four dimensional space, and are used in gyro-compasses. If it were not for the spatial properties of quaternions men would not have stood on the moon, and aircraft flight would be more dangerous than it is. Complex numbers fill out a two dimensional space, and are used to explain all electro-magnetic phenomena. Without the spatial properties of complex numbers we would live, very roughly, at a Victorian level of technology. Real numbers fill out a one dimensional space on the number line. Without real numbers we would live, very roughly, at an ancient Egyptian level of technology. Perhaps McGinn meant that "integers" do not have any spatial properties? This is conceivable at an ancient Egyptian level of mathematical sophistication, but in more mathematically advanced societies the integers are regarded as lying on the number line and can have spatial properties such as lying close, above, below, or between real numbers, complex numbers, quaternions, and octernions. Beyond the most elementary level of sophistication, numbers are spatial.

McGinn ends the paragraph by holding out the theoretical, but, in his opinion, unlikely possibility that there might be some spatial content to mind: "Spatial occupancy is not (at least on the face of it) the mind's preferred mode of being." He then continues.

> "Nor do we think of conscious states as occupying an unperceived space, as we think of the unobservable entities of physics. We have no conception of what it would even *be* to perceive them as spatial entities. God may see the elementary particles as arrayed in space, but even He does not perceive our conscious states as spatially defined – no more than He sees numbers as spatially defined. It is not that experiences have location, shape and dimensionality for eyes that are sharper than ours. Since they are non-spatial they are in principle unperceiveable."
>
> *Consciousness and Space* page 98.

McGinn might not have any conception of what it would be to perceive conscious states as spatial entities, but I do. I claim they are partial, bi-directional arrangements of perspexes. If God is the most imaginably supreme being then He is superior to me and can, at least, see conscious states as perspexes because I can. Alternatively, if God cannot see these things then I am mistaken in my ascription of what can constitute a conscious state. This might be, but if God does not see *numbers* as spatially defined, then mathematics is seriously mistaken. It is then a miracle, not an act of technology, that men walked on the moon, that television companies broadcast game shows, and that Apollo's stone alter can, theoretically, be doubled in size. Frankly, I do not understand McGinn's conception of space.

> "I am now in a position to state the main thesis of this paper: in order to solve the mind-body problem we need, at a minimum, a new conception of space. We need a conceptual breakthrough in the way we think about the medium in which material objects exist, and hence in our conception of material objects themselves. That is the region in which our ignorance is focused: not in the details of neurophysiological activity but, more fundamentally, in how space is structured or constituted. That which we refer to when we use the word 'space' has a nature that is quite different from how we standardly conceive it to be; so different, indeed, that it is capable of 'containing' the non-spatial (as we now conceive it) phenomenon of consciousness. Things in space can generate consciousness only because those things are not, at some level, just how we conceive them to be; they harbour some hidden aspect or principle."
>
> *Consciousness and Space* page 103.

It seems to me, that the essence of McGinn's conclusion is that consciousness is non-spatial, and that some new property of geometry or physics is required to explain how non-spatial things can be derived from spatial things. This is wrong.

Firstly, integers are, in themselves, non-spatial, but they lie on the number line which is a one dimensional thing. Mathematicians have no difficulty separating integers from other kinds of numbers on the number line. Hence they have no difficulty ascribing spatial properties to integers.

Secondly, the states of a Turing machine's finite state machine are, in themselves, non-spatial, but they can all be written onto a one dimensional tape in a Universal Turing machine that simulates any particular Turing machine. Thus the non-spatial states of any Turing machine can be converted to and from spatial ones on a Turing tape.

Thirdly, all of the states of a Turing machine can be seen as integer numbered points in a spatial lattice embedded in the program space of a perspex machine. The non-spatial, integer calculations are just restricted version of continuous, spatial calculations. There is simply nothing to prevent a continuous, perspex machine from operating on integer numbered points, so there is nothing to explain in how a non-spatial thing can arise within a spatial thing.

Fourthly, imagine hanging a picture on the wall of your room. The room is three dimensional, but the picture is two dimensional. Imagine highlighting one, straight line on the picture. The highlighted line is one dimensional. Imagine that the picture is a pointillist one so that there is no continuous line of paint, just individual points. In themselves the points are non-spatial, they have no area or volume, they are not solid, but they are contained in a one, two, and three dimensional space. Where is the problem with representing non-spatial things in space?

Drawing all of this together, we can say that McGinn claims that conscious states are non-spatial, but the reasons quoted here are not sufficient to support his claim. McGinn's case would be strengthened by dropping his weak supporting arguments and making a direct appeal to the reader's introspection. Perhaps, McGinn has better arguments in his favour and could usefully deploy these? But even if we allow McGinn's claim that mind is non-spatial, there is no reason to believe that mind cannot be embedded in space. If it were generally true that non-spatial things cannot be embedded in space then there would be no scalar fields in quantum physics: temperature would not exist, pressure would not exist, no images of any kind would exist. It would be impossible to see anything, including the words on this page.

My case is that non-spatial things are special cases of spatial ones, and that mental states are spatial. Specifically, that conscious states can be partial, bi-directional arrangements of perspexes. I try to strengthen my case in later chapters by giving philosophical arguments concerning the detailed kinds of consciousness that might be obtained by perspexes. Perhaps I could strengthen my case further by attacking a greater number of opposing philosophical arguments? I examine Clarke's argument next. Certainly, my case would be considerably strengthened by a successful outcome to the scientific task of constructing a perspex android that is conscious in the way hypothesised. So my effort might be better spent on scientific work rather than on philosophical work.

Clarke: The Nonlocality of Mind

In criticising Clarke[5] I extend the same courtesies as I did to McGinn, but in criticising Clarke in his own terms, I must abandon "objective" reality and enter the debate in terms of subjective experience. In doing this I am immediately in an unreconcilable position with Clarke because my experience of my mind is unlike Clarke's account of what it is like to have a human mind. Granted that we are all human, there must be at least two different gradations of human mind which correspond, I suppose, to people who have a high degree of visualisation and those who have a low degree. I am in the high visualisation group, and I take Clarke to be in the low visualisation group. If we accept this difference is a difference of degree, not a difference of kind, then we may usefully engage in a dialectic.

> "Concerning the nature of mind, then, it is entirely possible that mind is a derivative concept, reducible to some sort of physical mechanism. If, however, I acknowledge that the existence of mind is the primary aspect of our experience, then it seems unnatural to derive mind from physics, because this would be to try to explain something obvious and immediate (mind) from something (physics) that is an indirect construction of mind. So for me it seems a more fruitful method not to derive mind from physics but to reconcile the experience of mind with the world description of physics."
>
> *The Nonlocality of Mind* page 166.

Thus Clarke sets out his territory. He acknowledges that mind might be physical, but he starts from the subjective experience of mind and then seeks to explain how a particular mind describes the physical universe. At the end of the argument he makes claims about the physical universe, completing his transition from subjective experience to "objective" reality.

> "This mental world, the world of our awareness, comprises a mass of different thoughts which can be segregated into various categories – percepts, feelings, judgements, volitions etc. I want to suggest that the majority of these have little to do with Euclidean space. Certainly Euclidean space has a role in visual and propriomotor percepts, and to some extent in hearing. (One of the triumphs of modern psychology is the unravelling of the contributions of these different senses to our spatial perception.) In addition, many of our feelings – of anger, fear and so on – have important links with parts of the body and hence indirectly with space; but it would be hard to claim that this was the aspect of them that was uppermost in normal awareness. Other thoughts in our mental world, I would claim, have no direct link with Euclidean space at all."
>
> *The Nonlocality of Mind* page 167.

Clarke acknowledges that vision, the perception and enacting of bodily motions, and hearing involve a perception of space. He acknowledges that anger and fear are linked to the body, but claims that these, or I suppose any spatial perceptions, are not uppermost in one's mind when experiencing anger or fear. Perhaps I experience another kind or degree of anger and fear from Clarke. I can only ever be angered by physical things: humans rank highest on my list, followed by absurd communications on paper or by email. I suppose I might be angered by a surreptitious dose of adrenaline, but I expect that only a physical thing would be the focus of my anger.

The same goes for fear. I am afraid only of physical things: imminent, violent death comes highest on my list, personal confrontation next. I suppose I might be made afraid by a surreptitious dose of adrenaline, but I expect my fear would be focused on a physical thing. I can see no reason to believe that I ever experience anything in a non-spatial way – my earlier caveat about written words, excepted.

> “Neither, however, are these thoughts linked to any generalised space; because there are no spatial concepts at all that apply to most of them. Except in a purely poetic way, would it make sense, for example, to say that my realisation that $\sqrt{2}$ is irrational is ‘between’ my feeling annoyed with my secretary and my hearing an indistinct roaring noise from the central heating? All that can be said is that these various thoughts are buzzing around with some kind of togetherness, but without any sort of betweenness, nearness, or any other spatial relation to each other.”
>
> *The Nonlocality of Mind* page 167.

There are a number of ways space might enter these examples. Firstly in the experiences themselves; I cannot be angry with my secretary without visualising her face, body, clothes, desk, the things she had done, or failed to do that make me angry. I cannot imagine $\sqrt{2}$ without visualising part of the number line. I cannot hear a roaring noise from the central heating without visualising the pipes, boiler, or exhaust vents that might make the noise. Perhaps Clarke can be angry in the abstract, or can think of a number without reference to other numbers, or hear a noise without wondering where it comes from, but I cannot. As for these experiences being “between” something, they can certainly be laid out on a time line. I might happen to realise that $\sqrt{2}$ is irrational after I am angry with my secretary and before I hear a roaring sound from the central heating. Alternatively, my feeling angry with my secretary might be more extreme than my feeling of mild surprise at the irrationality of $\sqrt{2}$ and my faint recognition that the central heating is still roaring. I have no trouble ordering these things in a list, or assigning a degree of attention, annoyance, or whatever, to them by marking off my feelings on a continuous line. Such reportage is quite common in certain areas of experimental psychology and, whilst subjects are not particularly reliable in such reports, they find no particular difficulty in making reports in this spatial form. Clarke might not have any sense of space in these things, but I do, and so, it would seem, do most human subjects in psychological experiments.

> "To conclude this section I should return to the seminal work of S. Alexander (1920). He argued that all experience is spatial, even if only vaguely so. There was thus a mental space, as well as a physical space. The distinction between the two was the distinction of subject and object already referred to: mental space is enjoyed while physical space is contemplated. Yet enjoyed space and contemplated space, he claimed, match. There is a precise correspondence between the space of experience and the space of the physicist."
>
> *The Nonlocality of Mind* page 168.

My mental world seems to agree much more with Alexander's, than Clarke's.

> "Reading Alexander, one observes that all his examples of thoughts are visual ones. Here is a writer who is dominated by visual images and who therefore ignores the pre-eminence of non-spatial thoughts in our awareness. As a result he is lead to postulate two spaces with a correspondences between them. A fuller appreciation of the range of our thoughts shows that space only comes from percepts. If we take this into account the alternative picture arises in which enjoyed space is actually derivative from contemplated space."
>
> *The Nonlocality of Mind* pages 168-169.

Thus Clarke claims that the perception of space arises from our contact with physical space, but he claims that most of our experiences are non-spatial. However, I still do not know what non-spatial experiences Clarke has in mind.

> "While probably no one today would hold to this original Cartesian form, it is worth examining because it reveals many misunderstandings that beset other dualistic approaches. First, there is no reason to suppose that since Soul is non-spatial it is without parts. The analogy of a computer program shows that it may be appropriate to analyse a system into functional parts without these parts being spatially located. Second, it ignores the possibility of causal influences that act on a physical material system in a distributed way, such as occurs with superconductivity in which the charge carriers are distributed over a large region and the mechanisms governing their behaviour are global in nature."
>
> *The Nonlocality of Mind* page 170.

I am willing to accept that a soul might be functionally complex, but Clarke's claim that a computer program is non-spatial is wrong. Computer programs are laid out in a computer file with each symbol in a strict, linear sequence. Similarly, programs in a Universal Turing machine are laid out in a linear sequence on a Turing tape. The functional blocks of a program are tied directly to the text in a statically bound programming language, and are tied to the temporal sequence of a particular execution of a program in a dynamically bound language. Thus, the functional blocks are bound to spatially arranged blocks of program text, or else to spatially arranged stack frames in the call stack of a computer. I do not know of any non-spatial computer language, and I teach compiler theory and practice to computer science students, and have worked on British Standards Institute panels formalising computer languages. If a non-spatial computer language were to exist, I cannot imagine how any human could write programs in it except by considering linear sequences of things to be done, or linear sequences of facts to be brought into play in a proof path. Clarke might be able to imagine a non-spatial computer language that can be programmed non-spatially, but I cannot, and nor, it would seem, can any computer scientist.

Clarke's claim that Descartes did not consider that mind might be spread everywhere in space is literally (textually) true, but what of it? Being located everywhere in space is just as much a spatial property as being located at one point in space or at an isolated neighbourhood in space. The nonlocality of mind does not deny the influence of space, on the contrary, it asserts that the whole of space has an influence on mind.

> "The way ahead, I believe, is to place mind first as *the* key aspect of the universe. It is what determines the operator-histories without which the universe does not exist. We have to start exploring how we can talk about mind in terms of a quantum picture which takes seriously the fundamental place of self-observation; of the quantum logic of actual observables being itself determined by the current situation. Only then will we be able to make a genuine bridge between physics and psychology."
>
> *The Nonlocality of Mind* page 175.

Clarke claims that we need a radical change in our understanding of the physics of the universe if we are to understand how psychological states, such as consciousness, enter the physical universe. By contrast, I describe how partial, bi-directional arrangements of perspex neurons might supply these conscious psychological states. In this, I draw on my mathematical unification of the Turing machine and projective geometry[4]. I do not need to change the geometry by which we imagine the universe operates to do these things but I could change the geometry of the perspex machine to match any given geometrical account of physical space.

It is difficult to see how to strengthen Clarke's position as set out here. It seems to amount to nothing more than the bald claim that the majority of human experience is non-spatial. Perhaps Clarke can find better examples, or show where I am wrong in supposing that the examples he gives do involve space? Until this is done, there is no point of contact between Clarke and me. It appears that we have different kinds or degrees of mind.

Conclusion

McGinn and Clarke appear to have a different conception of number from mine. Perhaps they regard numbers as Platonic ideals that exist in a non-physical world of ideas. I take the materialistic line that numbers are particular, physical, relationships between physical things. For me, marks on paper can be numbers; electromagnetic states of a digital computer can be numbers; and, I suppose, certain states of my neurons can be numbers. For me, the symbol (x, y) is a complex number and can be subject to physical, arithmetical operations. The symbol (x, y) is also the Cartesian co-ordinates of a 2D point and can be subject to physical, spatial operations. I cannot see how the complex number (x, y) can fail to be the spatial, Cartesian co-ordinates (x, y) because all of the properties of Cartesian co-ordinates are present in the number. Given any complex number one can plot its position in Cartesian co-ordinates. Seen in this way, complex numbers have spatial properties. Conversely, 2D Cartesian co-ordinates can be written down as complex numbers, so points in 2D space have numerical properties. I can construct similar arguments with 1D real numbers, 4D quaternions, and 8D octernions, but this would be too much of a digression. I am at a loss to know how McGinn and Clarke think of numbers and space, or why they think that the supposed non-spatiality of mind is any impediment to embedding mind in space. Why change the laws of physics, when we can simply ignore the, supposedly, unwanted spatial properties of the universe? After all, digital computers since the time of Babbage have ignored the apparent continuity of space and imposed an integer numbering on it. Babbage did not have to change the laws of physics to create a digital computer, so why should we have to change the laws of physics to create a conscious computer?

What I find personally disturbing about McGinn's and Clarke's arguments is that they assert that the majority of human experiences are non-spatial. If so, it appears to me that I am not privy to the majority of human experiences. But I like being able to think in pictures, sculptures, motions, voluntary actions, and words. I like the fact that words are tied to spatial things and mark the boundaries of my non-verbal thoughts. Now that I can manipulate written words like a regular member of the human race, I would feel a tragic loss if any of these spatial experiences were taken away from me. I do not think I would like any non-spatial experiences because, I imagine, they would not cohere with the experiences I am accustomed to and like. I do no like incoherence in my experiences. I do not want the non-spatial sort of consciousness McGinn and Clarke want to offer. I would feel diminished by it.

McGinn and Clarke offer a council of despair in which an understanding of consciousness is beyond our current understanding. But it seems to me that I do understand consciousness in terms of a partial, bi-directional relationship between perspexes. If I am right in this, then I can restrict this spatial form of consciousness to the non-spatial, verbal sort that McGinn and Clarke seem to want. If I am right, I can offer them what they want, but they can offer nothing.

On a personal note, I would say that dyslexia is a terrible affliction, but it can be overcome. If a high degree of visualisation can arise in the human mind only in response to dyslexia then I am content to suffer this specific learning difficulty. On the other hand, if a high degree of visualisation can be acquired without dyslexia then I have no objection to my kind of person being eradicated from the gene pool by genetic engineering, unless, of course, dyslexia is linked to some other beneficial trait for the individual or the population. I can well imagine that mental abilities vary in kind as well as degree amongst the human population so that the whole population is more resilient to challenges if all kinds of people are allowed to come into existence and participate in society. I believe that the careful introduction of androids into society would strengthen society and increase our prospects for survival.

Questions

1. What non-spatial experiences do people have?
2. How can numbers be considered to lack spatial properties?
3. Are there any properties of consciousness that are not in a partial, bi-directional relationship?
4. Is there any property or kind of consciousness missing from the account of consciousness given in this book?
5. What kind of consciousness do you want to have explained?
6. What kind of consciousness would you like to have?
7. Should an android be allowed to change the kind or degree of consciousness it possesses? If it does this, what moral responsibility should it bear for its choices and the actions it performs in each of these states of consciousness? Who, or what, could judge such an android and bring legal sanctions to bear?
8. Should legal issues dominate moral issues when regulating a human, android, or mixed society?
9. Should an android be allowed to try out different kinds or degrees of consciousness and then be permitted to try to promote its chosen kinds or degrees of consciousness in the human race by genetic engineering?

Further Reading

1. **Alexander, S.** *Space, Time, Deity* Macmillan, (1920).
2. **Anderson, J.A.D.W.** "Visual Conviction" *Proceedings of the fifth Alvey Vision Conference*, University of Reading, England, pp. 301-303, (1989).
3. **Anderson, J.A.D.W.** "Representing Visual Knowledge" in *Philosophical Transactions of the Royal Society* vol. 352, pp. 1129-1139, (1997).
4. **Anderson, J.A.D.W.** "Perspex Machine" in *Vision Geometry X1*, Longin Jan Lateki, David M. Mount, Angela Y. Wu, Editors, *Proceedings of SPIE* Vol. 4794, 10-21, (2002).
5. **Clarke, C.J.S.** *"The Nonlocality of Mind"* in **Shear, J.,** editor, *Explaining Consciousness – 'The Hard Problem'* MIT Press, pages 165-175, (1995).
6. *Collins English Dictionary* Harper Collins, (1991).
7. **McGinn, C.** "Consciousness and Space" in **Shear, J.,** editor, *Explaining Consciousness – 'The Hard Problem',* MIT Press, pages 97-108 (1995).

8. **McGinn, C.** *The Problem of Consciousness* Basil Blackwell Ltd., (1991).

9. **Noë, A. & Thompson, E.** *Vision and Mind: Selected Readings in the Philosophy of Perception* MIT Press, (2002).

10. **Popper, K.R. & Eccles, J.C.** *The Self and its Brain - an argument for Interactionism* Routledge, (1995). First published 1977.

CHAPTER 8

Free Will

$$\text{jump}(\vec{z}_{11}, t)$$

Introduction

Philosophers have debated the nature of free will for centuries; some even believe that all of the interesting questions about free will have been asked, leaving only psychological questions of what it like to experience free will unasked. If this is correct, then I can simply pick a philosopher to stand in for the philosophical arguments I might make. Daniel Dennett[4] will do:

> "What we want when we want free will is the power to decide our courses of action, and to decide them wisely, in the light of our expectations and desires. We want to be in control of ourselves, and not under the control of others. We want to be agents, capable of initiating, and taking responsibility for, projects and deeds. All this is ours, I have tried to show, as a natural product of our biological endowment, extended and enhanced by our initiation into society."
>
> *Elbow Room* page 169.

Thanks, Dan. That leaves me to propose how perspexes might give a robot will; how the design of a robot's body and the initiation of a robot into society might, in general, ensure that its will is free of direct control by others and free of "dehumanising" physical and social constraints; and how perspexes might give rise to consciousness of free will.

Will

We could argue about the nature of *will* for a long time, and thoroughly consider the enormous philosophical literature on this subject, or we could look up "will" in a dictionary[5]. Will is, "the faculty of conscious and deliberate choice of action." Previous chapters show how to implement all of these elements in a robot, but it might help to make this explicit here, before we consider what more needs to be done to give a robot *free will*.

Firstly, let us define that a *motion* is $\vec{x}\vec{y} \to \vec{z}$. Hence, within a perspex program, a motion is a particular multiplication of matrices $\vec{x}$ and $\vec{y}$ where the result is normalised and written into a matrix $\vec{z}$. See the introductory chapters *Perspex Matrix* and *Perspex Instruction*. Looked at in another way, a motion is the receiving, by the neuron defining the motion, of two afferent signals from the neurons $\vec{x}$ and $\vec{y}$, followed by the sending of the combination of these signals, as an efferent signal, to the neuron $\vec{z}$. See the chapter *Perspex Neuron*. Looked at in another way, a motion is a geometrical, perspective motion $\vec{x}$, of the primitive, geometrical object $\vec{y}$, or *vice versa*. See the chapters *Perspex Matrix* and *Perspex Simplex*. Thus "motion" has its common meaning of being the movement of an object in space, as well as being an abstract, but physically real and specific, motion in a neural network and in a program.

Secondly, let us define that a *selection* is $\mathrm{jump}(\vec{z}_{11}, t)$. Hence, within a perspex program, a selection is a particular decision to do one of a collection of four things t, depending on whether the number $\vec{z}_{11}$ has a value that is less than zero, equal to zero, greater than zero, or else equal to nullity. See the chapter *Perspex Instruction*. Looked at in another way, a selection is the sending of a transferent signal from the neuron defining the jump to one of four neurons specified by t_1, t_2, t_3, and t_4. See the chapter *Perspex Neuron*.

A selection can also be seen in an infinite number of geometrical ways, but it seems sensible to employ a way that supports the sensorimotor sentences in a robot's internal language, as introduced in the chapter *Beyond Language*.

Imagine that a robot touches or looks at part of a geometrical object described by the simplex $\vec{x}$, and then its hands or eyes make a sudden movement, or saccade, so that it touches or looks at a part described by the simplex $\vec{z}$. If $\vec{x}$ and $\vec{z}$ have a volume, so that they are non-singular, then we can compute $\vec{y}$ using a matrix inverse. On the other hand[1], if either or both are singular, we could project them into a lower dimension and try to compute the inverse again, and so on, until we arrive at a non-trivial solution, or else to nullity. There are very many generalised inverses we could use when a simplex is singular[1]. Having obtained $\vec{x}$, $\vec{y}$, and $\vec{z}$ from a geometrical object, described as perspexes, we need to store the perspexes somewhere in program space.

There is an infinite number of ways we could store the perspexes, but if we store them in an identity perspex machine[2], which is a very simple sort of perspex machine, each triplet of $\vec{x}$, $\vec{y}$, and $\vec{z}$ perspexes will be stored in one 3D subspace of program space, fixed at an integer-numbered time. The $\vec{x}$ perspex will be stored at the distance one along the *x*-axis, $\vec{y}$ at the distance one along the *y*-axis, and $\vec{z}$ at the distance one along the z-axis; and the identity matrix will be stored at the origin of each 3D subspace. With this configuration of program space, a perspex machine steps through successive 3D subspaces in temporal order, executing $\vec{x}\vec{y} \to \vec{z}$ as it goes, and always jumping forward in time to the next 3D subspace. Thus, wherever and whenever the triplet of perspexes was touched, seen, or imagined, it will end up in a linear, temporal sequence. Furthermore, each triplet in this sequence has the grammar $\vec{x}\vec{y} \to \vec{z}$ where $\vec{x}$ is an object, or noun, $\vec{y}$ is a motion, or verb, and $\vec{z}$ is an object or noun. Thus the sequence of perspex triplets forms an alternating sequence of sensorimotor nouns and verbs. A slightly more complicated scheme sends the resultant perspex $\vec{z}$ into the position of $\vec{x}$ in the immediately future 3D subspace, so that each subspace contains a non-verb couplet, exactly like the sensorimotor sentences in the chapter *Beyond Language*.

If we apply this way of looking at objects everywhere, then all sensations are recorded as sensorimotor sentences with the grammar $\vec{x}\vec{y} \to \vec{z}$. In a practical robot, with a limited memory, the sentence terminates when all of the pre-allocated identity matrices are used up and the perspex program encounters nullity, which is a halting instruction, at the origin of the next 3D subspace. This termination of a sentence is an inherent property of a practical robot with limited memory and can be seen as a limited temporal attention span, or as a limited length or perspex neural fibre providing the thread that ties the nouns and verbs together. Alternatively, the sentence will be terminated when the android fails to touch or look at something, in which case it introduces nullity as a sensory perspex $\vec{x}$, $\vec{y}$, or $\vec{z}$. This termination can be seen as arising causally from the termination of the sensory input.

It is technically easy to arrange that all sensory inputs are treated in this way, but if we do this, we provide a robot with a language faculty in each sensory modality. The sensory sentences might contain only Turing-computable values, in which case the sentences are linguistic sentences of symbols; but if they contain only Turing-incomputable values then the sentences retain the same grammatical structure as a linguistic sentence, but with Turing-incomputable values taking the place of symbols. A perspex machine can operate symbolically and/or in the continuum.

A very significant side-effect of supplying such a linguistic faculty is that all sensory inputs are organised in a temporal sequence, so that a robot will experience the world as an ever changing 3D space, just like us. Furthermore, it will have the potential to relate all of its sensory inputs to words; but some inputs, or feelings, might be ineffable, in the sense that they are Turing-incomputable, and therefore cannot be put exactly into any symbolic form.

A further consequence of this linguistic organisation of sensations is that all sensations are broken up into distinct simplexes, in fact, triplets or couplets of simplexes, even if the sensed objects are continuous in the world. Thus a robot will sense a world of discrete objects, not an undifferentiated continuum. In order to think in the continuum it must either access the continuum of numerical values directly, or else engage in visual and not linguistic reasoning. A robot can experience a range of mental styles from a predominantly linguistic one, to a predominantly visual one, just like us.

Finally, let us define that an *action* is the whole of the perspex instruction: $\vec{x}\vec{y} \to \vec{z}$; $\text{jump}(\vec{z}_{11}, t)$. As we have just seen, an action can be a physical action in the world, a single instruction in a perspex program, the transfer of signals into and out of a single neuron, or a single phrase of a sensorimotor sentence. All of these forms have an identical structure. All of the forms internal to a robot can be caused by physical actions in the world. Conversely, a linguistic sentence, a neural signal, and a program instruction, can all cause physical actions in the world. There is no mystery to how sentences arise from sensations or how sentences cause a robot's body to move. Perspexes provide a simple, causal physics for arranging these things.

Now that we have dealt with the geometry, let us play a word game to see how to provide a robot with will. Will is[5], "the faculty of conscious and deliberate choice of action."

We suppose that a "faculty" is the computational machinery needed to do some mental thing; but this is only to say that the thing is provided by a perspex machine so we may drop the word "faculty" from a perspex definition of will, leaving: will in a perspex machine is "conscious and deliberate choice of action."

What is the import of the word "deliberate" in this definition? If Mr. Gerald Mogglington-Moggs has a night mare that causes him to jump up and knock over a coffee table, including a cup of coffee, then we would not hold him responsible for his actions, even if he had been dreaming that he was under a physical attack and that his only chance to survive was to jump up and knock over the coffee table, thereby frightening his attacker away. Having done this, and woken up, he would see that his attacker was no longer there. In this case all of the actions are planned, and carried out successfully; but we do not hold Mr. Mogglington-Moggs responsible for the stained carpet, because it was an accident of neurophysiology, not a deliberate act, that caused the dreamed actions to turn into bodily actions. In other circumstances we might hold Mr. Mogglington-Moggs responsible for his actions. Let us say, then, that if we find that actions are deliberate we hold a being responsible for an act of *moral will*, but if the actions are not deliberate they are an act of *will.* Thus we have that moral will in a perspex machine is "conscious and deliberate choice of action," but that will in a perspex machine is "conscious choice of action." By using this latter definition of *will* to arrive at *free will,* we make a simplification that allows us to reduce free will to a fairly simple arrangement of perspexes. It is far beyond the current state of mathematics to indicate what a definition of "deliberate" would be like in perspexes, so there is no immediate prospect of providing a perspex theory of morality.

Given this caveat about moral will, we press on. We have already defined "action." If we take a "choice" to be a "selection" then we have already defined "choice." A suitable sensorimotor sentence will describe the choice to do any particular thing, though, in general, such sentences will be more complicated than those above, in that they will allow a branching into one of four actions in each time step. In order to make this choice visually conscious we must arrange that the sensorimotor sentence is in a partial, bi-direction relationship with an image. Images of the predicted outcomes of the selections will suffice. Thus we arrive at a "visually conscious choice of action" or, by synaesthesia, to a "conscious choice of action" which is the desired definition of will.

Freedom

Freedom comes in many sorts. There is *freedom from* things – like oppression, fear, or compulsion – and *freedom to* do things – like eat walnut cake, escape an attacker, or carry out science disparaged by bureaucrats.

For an android we can arrange a degree of freedom from oppression by legislating for equality of treatment with humans and educating everyone to respect the place of androids in society. Alternatively, androids might take themselves off to remote parts of space, travelling faster than humanity can keep up with them; thereby ensuring that they are forever beyond the reach of human oppression. Freedom from fear might be arrange similarly, by outlawing the things that cause fear in an android; but we are getting ahead of ourselves, if we could give robots a feeling of fear that would be a spectacular achievement. Freedom from external compulsion might be similarly legislated, but freedom from internal compulsion, from their own programs, must be supplied in a different way. For that, they need free will. Freedom to do things is similar. There are no laws that prohibit the eating of walnut cake. Physical circumstances might allow us to escape an attacker. We might exercise free will and carry out some action that is disparaged and subject to external, oppressive, coercion despite the consequences.

Free will is important to us, and we can give it to robots.

Free Will

We have derived a definition of will. Will is a "conscious choice of action." For example, it is a partial, bi-directional relationship between, on one side, a sensorimotor sentence setting out choices and, on the other side, images of the expected consequences of those choices. Philosophers might not like this definition, but it is a good enough starting point for an engineer to start building a robot that might, one day, have free will like ours. Now suppose that a robot is programmed to acquire sensorimotor sentences from the world and to execute those sentences as programs. Each robot will see the world from a physically different view point, so it will acquire different sensorimotor sentences and will, therefore, execute different programs. The perspex brains of robots will depart from their original programs so that robots develop an individual nature which, of itself, will make robots both personally and historically creative. This individual and creative departure will be beyond any human power to predict. It will be beyond our power to arrange the world so as to coerce a robot, unless we come to know its individual nature in great detail. We can program a faculty of free will into a robot, simply by having it execute sensorimotor sentences as programs, but we would want more of free will for ourselves.

Paraphrasing Dennett, "we want the power, not only, to decide our courses of action, but to decide them wisely, in the light of our expectations and desires." If we are to give a robot this kind of free will, we must supply it with sufficient intelligence to have expectations and sufficient feelings to have desires. These things are discussed in the chapters *Intelligence* and *Feeling*. "We want to be in control of ourselves, and not under the control of others." Freedom from control by others is easily provided by programming a robot to execute some of its sensorimotor sentences. By providing intelligence we can hope to moderate these actions so that they are coherent with a robot's other programs, giving it an identifiable character, or self which can prevent it from executing actions that it can predict are undesirable. "We want to be agents, capable of initiating, and taking responsibility for, projects and deeds." Providing the perspex definitions to be an agent is easy, providing enough of these to plan a project is hard, but providing definitions capable of supporting responsibility, in the moral sense of "responsibility," is entirely beyond our current abilities. "All this is ours, I have tried to show, as a natural product of our biological endowment, extended and enhanced by our initiation into society." That may be true of humans, but it will take millennia of research and development to achieve a similar level of free will in a robot.

We might want all manner of kinds of free will for robots, but the practical requirements of science mean that we must start with the simple sorts, whatever these are. The execution of sensorimotor sentences is about as much as we might hope for in the immediate future.

Unfortunately, providing robots with free will, will mean that scientific experiments with them will be unrepeatable, at least to the extent that experiments with animals are unrepeatable. Quite possibly, robots will exercise free will in insane ways, so that their minds become incoherent. The challenge of providing intelligence for robots is to ensure coherency in a perspex mind that makes its interactions with the world useful. In this, sensorimotor sentences will play a key role.

Dissolving the Problem of Free Will

Philosophers have traditionally worried over the questions of whether free will and responsibility are compatible with determinism but I argue that free will and responsibility are independent of determinism so the philosophical problem of free will is dissolved. Let us take this philosophical problem in two parts, before considering the theological problem of free will.

First, suppose that the universe is deterministic, then a being's self is part of the deterministic web of the universe and any choices it makes are deterministic. The being may have free will in the sense, set out here, of making conscious selections, but cannot have free will in any non-deterministic sense because the non-deterministic mechanism of choice would, by hypothesis, lie outside the universe and hence outside the being's self. A deterministic being may be held physically responsible for all of the consequences of its actions, personally responsible for all of the consequences of its actions that it foresaw, and morally responsible to the extent that it acted according to acceptable codes of conduct. In these cases, punishment, reward, and treatment are deterministically effective in modifying a being's behaviour, whether they are self-administered or administered by an external agency. It cannot be the case that responsibility relies on any non-deterministic mechanism because then a deterministic being has no access to a mechanism for assessing its own responsibility or the responsibility of others. In these senses then, free will and responsibility are compatible with determinism.

Second, suppose that the universe is fully or partially non-deterministic, then a being's self is part of the stochastic web of the universe and any choices it makes are stochastic. The being may have free will in the sense, set out here, of making conscious selections, but cannot have free will in any non-stochastic sense because the non-stochastic mechanism of choice would, by hypothesis, lie outside the universe and hence outside the being's self. A stochastic being may be held physically responsible for all of the consequences of its actions, personally responsible for all of the consequences of its actions that it foresaw, and morally responsible to the extent that it acted according to acceptable codes of conduct. In these cases, punishment, reward, and treatment are stochastically effective in modifying a being's behaviour, whether they are self-administered or administered by an external agency. It cannot be the case that responsibility relies on any non-stochastic mechanism, because then a stochastic being has no access to a mechanism for assessing its own responsibility or the responsibility of others. In these senses then, free will and responsibility are compatible with a stochastic universe.

Taking these two cases together, we see that free will and responsibility are compatible with the causality of the universe they are embedded in, and it is (self-evidently) inconsistent to suppose that free will or responsibility rely on any mechanisms outside the causality of that universe.

This resolution of the problem of free will might not make philosophers happy, but it provides a perfectly general basis for the robot designer to implement free will and a sense of responsibility in a robot. The precise physics of the universe does not matter, the proposals made here can be put to the test in any universe.

Some philosophers argue that God cannot have free will because His perfectly good nature leaves Him no choice but to do the best thing. Such arguments are bunkum. If there are two or more equally good things to do then God requires a mechanism of choice to decide which one of the equally good things to do in any particular universe. Thus, even an omniscient and omnipotent God retains free will.

Some philosophers argue that God is the only agent in the universe because he determined everything. This is not necessarily so. God might have envisaged two or more equally good universes and chosen to create one of these using a perfectly random mechanism. Then God determined that there should be a universe, but did not determine which universe should come into existence. Thus, He did not determine the agents in this universe. But this is to say that He gave the agents in this universe free will with respect to Himself. The selections an agent makes are not determined by God, even if the universe has had a deterministic causality since its creation. If God is omniscient and can foresee everything that will happen in this universe then it is perfectly responsible of Him to order, at the point of creation, which souls should be raised up to heaven, and which should be cast into hell, even though those souls have not yet acted in a way to merit reward or punishment. Theologians might not like it, but this resolution of the problem of divine free will provides a mechanism by which God may make any agents in the universe subject to His judgement whether He created them directly, or whether they arose during the life of the universe. Robots and animals yet to come can be subject to God's judgement, just as much as humans, because any agent is capable of bearing responsibility for its actions.

Given the passion which surrounds theology, it is important to get clear what is being said here. I have said that it is conceivable that God envisaged two equally good universes. Envisaging two equally good, and mutually exclusive, things is sufficient to require God to have free will to decide which one to allow. He might have envisaged two such universes in order to require Himself to have free will. Having envisaged at least two universes he created one of them in a way that allowed all of the agents in that universe to retain free will with respect to Himself. Thus, both God and his creation retain free will. In this case, it is up to us whether or not we create robots, and whether or not we give them free will; however we decide, we bear the responsibility for that decision.

Questions

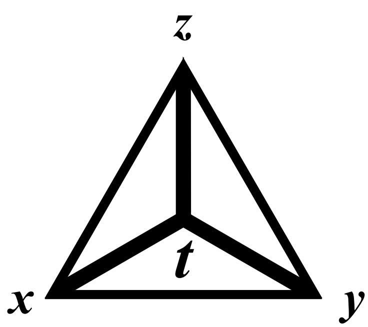
1. What kinds of consciousness of choices might we want a robot to have?
2. How could we take sanctions against a robot unless it has feelings and intelligence? In the absence of sanctions, how else might we protect ourselves from robots?
3. How can robots take sanctions against us, unless they are sufficiently like us to understand our feelings and motivations? How else might they protect themselves from us?
4. What would we want from a mixed society of animals, including genetically modified and wild humans, and robots? What might they want from such a society?
5. Are there any moral or scriptural prohibitions against giving robots free will?
6. Would giving robots free will help us to understand our own free will?

Further Reading

1. **Anderson, J.A.D.W.** "Representing Visual Knowledge" in *Philosophical Transactions of the Royal Society* vol. 352, pp. 1129-1139, (1997).
2. **Anderson, J.A.D.W.** "Perspex Machine" in *Vision Geometry X1*, Longin Jan Lateki, David M. Mount, Angela Y. Wu, Editors, *Proceedings of SPIE* Vol. 4794, 10-21, (2002).
3. **Anderson, J.A.D.W.** "Robot Free Will" in *ECAI 2002 Proceedings of the 15th European Conference on Artificial Intelligence,* Lyon, France, ed. F. van Harmelan, pp. 559-563, (2002).
4. **Dennett, D.C.** *Elbow Room* Oxford University Press, (2002). First printed 1984.
5. *Collins English Dictionary* Harper Collins, (1991).

CHAPTER 9

Intelligence



Introduction

Many, perhaps most, readers of this book will have taken part in an Intelligence Quotient (IQ) test at some time or another. These tests are all similar: find the pattern in this list of numbers and calculate the next one; find the pattern in this list of words and say which is the odd one out; or find the pattern of guests at a dinner table and indicate whether Albert or Beatrice is playing footsie with Charlie. These tests might be useful for organising teaching, but none asks the really big question: what is the pattern of everything – what is intelligence?

Intelligence has a lot of work to do. Firstly, intelligence must relate the world to ideas so that a being can see patterns in the world that afford it opportunities for action. Secondly, intelligence must relate ideas to the world so that a being can carry out actions that are meaningful in terms of its bodily needs, mental capacities, and social situation. This bi-directional relationship is reminiscent of visual consciousness, but intelligence must do more than consciousness. Thirdly, we want intelligence to see patterns within ideas themselves so that a being can learn from its interactions with the world and with other beings. Fourthly, intelligence must help the being stay alive by causing its body to act in one, coherent, way when danger or opportunities are present. This coherent kind of intelligence, I argue, leads to self-consciousness. Finally, intelligence must be astonishingly versatile.

I propose that intelligence is perspex symmetry, and that this manifests, to a greater or lesser extent, in the world, in a perspex brain, in actions, in communication, and in all manner of things that can be described by perspex programs. When we see symmetry between a complex program and the behaviour of the entire universe, or part of it, such as an animal or a robot, we have reason to believe that the behaviour is at least as intelligent as the program. Perspex symmetry gives us a reason for believing that things other than ourselves are intelligent. It explains, too, why we anthropomorphise things, seeing human intelligence in non-human things.

Symmetry is not a simple idea. Mathematicians define symmetry as being the same shape after some function has been applied, ostensibly, to change the shape in some way: rotations give rise to rotational symmetries, translations of the origin give rise to translational symmetries, and binary reflections in an axis give rise to reflectional symmetry. The affine and perspective symmetries provide many, if not all, of the symmetries we humans can see. All of the rational numbered symmetries are computable, and some of the real numbered symmetries are computable, but there are other kinds of symmetry. There are Turing-incomputable symmetries, such as rotation by an incomputable angle, translation by an incomputable length, and reflection in an incomputable axis. These Turing-incomputable symmetries provide a kind of intelligence that cannot be achieved by any digital computer, and cannot be put, exactly, into words, but a perspex machine can have these kinds of intelligence. A perspex machine can exhibit kinds of intelligence that cannot be tested in any written IQ test. It can have degrees of intelligence that cannot be measured by any kind of quotient. Quotients are rational numbers, but the degree of intelligence exhibited by a perspex machine could be a Turing-incomputable real number. A perspex machine can be too clever for words. It can be too clever to have an IQ score.

If we humans are limited to verbal kinds of intelligence, then perspex robots might be intelligent in ways we cannot recognise. Far from language being the hall mark of intelligence, non-linguistic, perspex robots might pass beyond the reach of our philosophy and understanding. This might be wonderful from an ecological point of view, but it would be a methodological disaster in the early stages of research. It would be easier, and safer, to restrict perspex robots to the Turing-computable, rational, affine and perspective symmetries, before giving them access to modes of thinking that might be inherently superior to our own.

Conscious Intentionality

In philosophy “intentional” means that a mental state refers to the world. So, for example, the concept *green* is intentional when it refers to the stems and leaves of flowers in a garden (see the chapter *Visual Consciousness*). We have already seen how sensorimotor sentences relate from the world to a mind, and *vice versa*, in the chapters *Beyond Language* and *Free Will*. When sentences in a sensory cortex have a partial, bi-directional relationship with sentences in a motor cortex this relationship is sufficient to make both sentences visually conscious and hence, by synaesthesia, to make them conscious. Of course, this is a very thin sort of consciousness, and a very thin sort of intentionality. We would want more for ourselves and for robots.

We want to have intentions in the more common, and specific, sense of having purposes in mind for our actions; actions that satisfy our various bodily, emotional, and intellectual needs. If a robot is to have intentions of this sort, it is going to have to have these various kinds of needs, and quite a lot of intelligence to understand the world and itself well enough to have purposes. We will discuss bodily and emotional needs in the chapter *Feeling*, here we concentrate on intellectual needs and purposes.

The most fundamental, intellectual need of a robot in possession of free will, as defined in the chapter *Free Will*, is to have the wherewithal to prevent itself from being driven insane by executing programs that are impressed on its mind by sensation. One way to do this is to arrange that sensations are presented in a form that is inherently meaningful and useful to a robot, so that the execution of these programs is inherently purposeful and not incoherent or insane. If we are to know what is useful to a robot we must have in mind some notion of the kind of intelligence it should have.

In the chapter *Perspex Matrix*, we discussed the importance of being able to undo operations in the world so that we can do them again. Operations like cutting food, eating, writing, knitting, sewing, hammering, walking, talking, breathing, yawning, blinking, winking, reading, shivering, having sex, falling asleep, waking up, and many others, all involve repeating an action. Repeating actions is one of the most fundamental things that animals do. In fact, almost nothing that an animal does is unique; almost everything can be done again. Killing oneself is a counter example – and not one to be recommended under ordinary circumstances!

A repeatable operation is a symmetry. For example, if we have our eyes open and performs the operation of blinking, this ends with our eyes open again. So the operation of blinking, which ostensibly changes the position of our eyelids, actually leaves our eyelids in the starting position – as required of a symmetry. In fact, the combination of any function and its inverse is a symmetry.

Mathematicians say that the combination of a function and its inverse is a trivial function called the *identity function*. It has no nett effect, so it does not change the identity of anything it is applied to in any way. However, blinking is a useful funtion. During the function of closing our eyelids, liquid is transferred from the tear ducts to the surface of the eyeball. Dirt and dust are transferred onto the edge of the eyelid where they can be washed away by tears, or else rubbed away by the hands. Having our eyes closed can be useful, even though it makes us temporarily blind.

Closed eyelids afford some protection to the optical surfaces of the eyeball; these surfaces do most of the focusing of light in the eye. Eyesight would be badly degraded by scratching them, as would happen by accident if the eyelids were always open. The inverse function of opening the eyelids also helps spread tears on the eyeball which, again, helps to clean and lubricate them. Taken together the function of closing the eyelids and its inverse, opening the eyelids, is a geometrical identity function with no nett effect, but it plays a vital role for an animal. The biologically important thing is that the function and/or its inverse has a useful side effect. In any case, it is unlikely that we will give up sex just because mathematicians say that, in general, coitus is a reversible function.

Despite the mathematician's objection that an identity symmetry is trivial it is useful to find out that a function has an inverse, because both the function and its inverse are needed to repeat an action. It is useful to find this symmetry so that we know how to repeat an action. This increases a being's ability to interact with the world in repeatable, that is, predictable ways. Symmetry helps a being stay alive and provides a lot of what we call intelligence. Perhaps it provides the whole of intelligence?

Some neurophysiologists have argued that animal brains compress sensory signals so that only the important details are recorded, and so that neural processing can be done in a small enough volume of brain to be carried around inside our heads! Mathematicians require that compression can be undone again. In fact, they go to some trouble to ensure that compression and decompression can be applied infinitely often so that, for example, the millionth copy of a CD sounds as good as the first. Thus, to a mathematician, compression is a repeatable operation – it is a symmetry. But this is not to say that the compression beloved of neurophysiologists is completely symmetrical.

Neurophysiologists do not want compression to be perfectly undoable. They do not want a one-to-one relationship, an isomorphism, between a sensation and its neural encoding. Neurophysiologists *do* want to loose sensory detail that is unimportant to an animal. They want everything that is in the neural encoding to be present in the world, but they do not care if the world contains unimportant variations. This relationship is called a homomorphism. It is vital that the world can vary in unimportant ways so that, for example: an animal can take another bite of food, despite the fact that the amount of food available is one mouthful less than it was on the previous bite; or so that an animal can breathe, despite being a few seconds older than the last time it took a breath; and so on. It is absolutely vital to animals that their minds can ignore unimportant detail. It is vital that their minds support homomorphisms.

Mathematicians, being logical folk, observe that a homomorphism could run the other way. A homomorphism could preserve all of the detail of the world in the neural encoding and ignore everything else that is encoded there such as motivations, desires, and so on. A homomorphism that runs both ways is an isomorphism so, to a mathematician, an isomorphism is a special kind of homomorphism. When I say it is important that animal minds contain homomorphisms, I do not preclude the possibility that they contain isomorphisms. Similarly, being a partial bi-directional relationship, or homomorphism, does not preclude being a total bi-directional relationship, or isomorphism. Visual consciousness can be homomorphic or isomorphic. However, when we look at the world all of the detail in the image on our retinas is in the world, but the world contains detail that we cannot see such as the actual size and distance of an object. We can see only the ratio of these two measurements, under normal circumstances. Normally, visual consciousness is a homomorphism.

Homomorphisms are symmetries. The detail that is preserved in an image on the retina can be projected back onto the world then back into the eye again, without making any difference to what is seen. This is what happens when a very good painter sees something in the world and paints a *tromp l'oeil* image of it on a wall; the wall then looks just like the thing the painter saw.

If we take neurophysiologists at their word, the encoding, or perception, of everything is a partial symmetry. Perception is, and supports, an awful lot of what a mind does, so symmetry provides an awful lot of intelligence.

Some psychologists have proposed that animals, particularly primates, see symmetries because symmetries reduce the amount of detail that needs to be encoded in a brain. Almost all biological things are bilaterally symmetrical, so a brain can record one half of an object and generate a picture or description of the other half when it is needed.

Symmetry is also important to the way an animal moves. In theory, knowing how our left leg walks means that we need only apply the same knowledge to our right leg. Knowing how to eat with our left hand means that, in theory, we need only apply this knowledge to our right hand in order to eat with it. Sadly, the left and right parts of our bodies are seldom as obliging as this, so that mammals fall into predominantly left and right handed classes, with very few, if any, totally ambidextrous individuals.

All of the above arguments show that symmetries are important to animals; so symmetry provides a large part of intelligence, if not all. Furthermore, the side-effect argument deployed in the discussion of identity functions can be extended to compute every mathematical thing. For example, searching all Turing-computable things for a symmetry allows us to do any Turing-computable thing *en passant.* More generally, searching all perspex computable things allows us to do all perspex computable things *en passant.* This is trivial, except in so far as it allows symmetry and, by my definition, intelligence to do anything at all, but there are important kinds of symmetry, as we have seen.

Symmetry is vital to the planning of acts. Imagine that you want to be rich, famous, or successful. Do you imagine having lots of money, adoring fans, or prizes? How do you imagine getting what you want? Do you imagine doing various things that get some money, or bring you to the attention of a few others, or win one prize? If you imagine this, and imagine repeating the process until you have enough of what you want, then that repeated action is a symmetry. On the other hand, if you imagine yourself now, and imagine yourself how you want to be the fact that you imagine yourself in both situations is a homomorphism and, therefore, a symmetry. All thinking about self involves symmetry, so any self-conscious act involves symmetry.

If a perspex android has any intentional thought then there is a homomorphism from the perspexes making up its thoughts to the perspexes making up the geometrical structure of things and their motions in the world. Thus mind includes part of the world "outside" a brain. Furthermore, intentionality implies symmetry, so even un-self-conscious purposes involve symmetry.

The search for symmetry can do any mathematical thing. So any action, whether purposeful or not, can be described by the search for symmetry. The search for symmetry can be the driving force of all Turing computations and super-Turing computations, but we do not currently know how to do everything we want for a robot. Symmetry is vitally important to an animal or robot, and makes perceptions coherent with the layout of things in the world, but it is not enough to guarantee that a robot will stay sane.

We might provide some guarantee of sanity by supplying robots with sane birth programs in read only memory, but this would guarantee the sanity only of what we supply: breathing, blinking, a faculty for language, or whatever. No doubt being sane involves controlling the amount of free will a robot has, and the boundaries of where it can apply free will, but these are issues that demand practical experiment. They are far beyond the reach of contemporary theory.

We might, however, supply sanity constraints by arranging that robots compete with each other, and with us, under the control of a genetic algorithm – a mathematical model of biological evolution. On the assumption that we are mostly sane, and that sanity increases survival, androids will have to match our sanity to survive under these conditions. Naturally we would want to arrange this competition in a way that is not too harmful to either party, as discussed in the chapter *Spirituality*.

Conscious Learning and Curiosity

Curiosity, famously, killed the cat. But the need to learn about its environs keeps a predator, like a cat, alive. Watching a cat hunt is quite amazing. It hides in long grass, or walks inside hedges, or along hidden branches then pounces on its prey. Cats, evidently, have some kind of knowledge of camouflage and stealth; they behave in purposeful ways that remind us of our own conscious actions. A feral cat that spends too much time hiding and stalking without killing prey will soon die of hunger; while a domestic cat would eat rather more of the food provided for it by its owners. So it would seem that a cat's actions are purposeful. They are, for example, to do with getting food, and are not random. Perhaps cats are not conscious of self, but they do seem to take a conscious interest in their environs, in prey, in other cats, and in non-prey animals. In all of these cases cats behave in a way that is usually appropriate to the purpose of keeping a cat alive and well. Mr. Mogglington-Moggs regularly eats birds, but he does not waste time, and risk injury, by trying to eat me!

Psychologists and philosophers argue that we can know very little of what goes on in an animal's mind, but with access to computers we can build mathematical models and practical examples of minds. We can know by analogy with perspex machines that if a cat has sensorimotor sentences in its sensory and motor cortexes, and these are related by a partial, bi-directional relationship then a cat is conscious of its actions. It is conscious of exploring, hiding, and pouncing. To the extent, if at all, that it can plan alternative approaches to a kill, it is conscious of a hunting strategy. Now, what will we have to do to provide this much for a robot?

Sensorimotor sentences just grow. They come into being as a causal consequence of sensing the world or moving one's body. In the chapter *Beyond Language*, I hypothesised that sensorimotor sentences and the relationships between them grow according to how regularly they are invoked by sensation and action, so that unused sentences and relationships die away. When a perspex machine learns anything, it does so by growing and/or killing perspexes. A perspex machine can visualise its own perspexes so it can become visually conscious of its learning, but what is the analogue of curiosity?

The search for symmetry can find relationships between sensorimotor sentences. The way a cup feels is the same, is symmetrical, on many occasions. The way "cup" sounds is the same, is symmetrical, on many occasions. A symmetry finding algorithm can find the homomorphism, that is the symmetry, between the occurrence of the sound "cup" and the feeling of a cup, if it can be found at all. If a robot can trigger a search for symmetry, in the way a cat can pounce, or can plan a search for symmetry, in the way a cat might be able to plan a hunt, then the intention to find symmetry is some kind of control sentence.

Any control sentence might be in a partial, bi-directional relationship with measures of the amount of computer memory or processing time needed and available, and the expected advantage of finding the symmetry. Thus, a robot can be visually conscious of a control sentence, or of any perspex program for that matter. If we define curiosity as searching, then we know what it means for a robot to be conscious of its curiosity as well as being conscious of its learning. It need not, however, be conscious of its self.

Consciousness of Self

We know what it would mean for a robot to be conscious of its body and the motions of its body; its neurons and the functioning of its neurons; its programs and the functioning of its programs; and of its sentences and their relationship to the world, including its body and brain. If we supply enough of these things then I suppose the robot can be conscious of self. If you deny this, then tell me what else you want of self and I will invoke a perspex sentence to supply it. By responding to your criticisms in this way we should, between us, eventually arrive at a complete description of the behaviour of a robot that is conscious of its self. But if a robot is to be conscious of its self, there must be some self to be conscious of.

We have seen that free will imposes individuality on a robot, so that its thoughts and behaviour are different from other robots. Individuality is an important part of self, as demonstrated by English sentences containing the phrase "myself." In the absence of individuality there would be no point in saying "my." The word "my" is only useful in contrast to words like "your", "our", and "their." But perhaps this linguistic insistence on individuality has more to do with the possession of a body than it has to do with distinguishing one's own mind from the minds of potential clones. Having a body means that a brain must decide to do one, gross, thing when threatened, or when opportunity presents itself. As we saw in the chapter *Visual Consciousness*, deciding to move either left or right can be important both to avoiding being eaten by a dangerous tiger, and to taking the opportunity to get some tiger steaks to eat. Just surviving in a body, means that there is a huge evolutionary pressure on developing some kind of knowledge of individuality; some knowledge of self and other.

If a robot lives in society with others, then it will have more chance of surviving if it can predict the behaviours of other individual robots and animals. An appreciation of self and other might evolve in competition with animals and other robots, but I do not know how to define it mathematically, or how to program it. The most I can certainly do, is provide a robot with individuality. If you accept my arguments about free will and consciousness then I can supply these things too. Now, let us turn to feelings.

Questions

1. How much of human intelligence can be usefully provided by the affine and/or perspective symmetries? Can humans see, touch, or otherwise feel any other kind of symmetry? If not, how do mathematicians understand other kinds of symmetry?
2. Is there any way to make an efficient computer program without making use of symmetry?
3. Is there any way to make an inefficient computer program by any means other than failing to make use of symmetry?
4. Is a computer program that exploits symmetry symmetrical? Alternatively, is symmetry of a program a sign of inefficient redundancy? What conclusion should we draw from the bi-lateral symmetry of animal brains?
5. Can counting be fully explained by physical symmetry with physical tokens, so that there is no need of Platonic integers that exist only in a non-physical mental space? Can measurement be similarly explained by symmetry with a standard length, so that there is no need of Platonic real, complex, quaternion, or octernion numbers? If so, can all concepts be explained by physical symmetries, so that there is no need of any non-physical, mental space?

Further Reading

1. **Dreyfus, H.L.** *What Computers Still Can't Do: a critique of artificial reason* MIT Press, (1997). First published 1972.
2. **Scruton, R.** *Animal Rights and Wrongs,* Demos, (1996).
3. **Shear, J.,** editor, *Explaining Consciousness – 'The Hard Problem'* MIT Press, (1995).

CHAPTER 10

Feeling

$\vec{x}$ $\vec{y}$

Introduction

Twinkle, twinkle, little bat!
How I wonder what you're at!
Up above the world you fly,
Like a teatray in the sky.

Alice in Wonderland[1] pages 97-98.

Read in one way, this is a nonsense poem by Lewis Carroll. But if we know that "Lewis Carroll" was a pseudonym of Charles Dodgson, and we know that Charles was a mathematics don at Oxford, and that one of his colleagues was nick-named, "The Bat," this poem takes on a quite different interpretation. We would then have to know quite a lot about the psychology of Charles Dodgson, The Bat and their professional relationship, to know whether the poem was intended as a complement or an insult.

Read in another way, the poem states that its author is wondering, or was wondering in 1865, about the motivations of a bat (a flying mammal of the order *Chiroptera*). Such creatures are unlike us in many ways, but it still seems to make sense to ask what it would feel like to be a bat.

Farrell[2] asked the same question in 1950; he wondered what it would be like to be an opium smoker, a bat, and a Martian. He said that scientists might discover everything about the behaviour and neurophysiology of a human or a bat, but that science has no way of knowing what experiences feel like to the being having them. However, he supposed that he could become an opium addict and know what that is like. Humans can find out what it feels like to be many kinds of people. As a child I wondered what it would be like to be a scientist, or to be a sub-aqua diver like Jaques Cousteau, or fly in space like Yuri Gagarin, or travel in time like Dr. Who. I became a scientist and I know what this is like in myriad detail. I became a sub-aqua diver who has been in France and spoken French, I know what this is like in myriad detail. I have flown on the edge of the atmosphere in Concorde, and been weightless underwater, and met Russian people, but I have never flown in space. I can imagine what it is like, based on these experiences, but unless I actually do it I suppose I will never fully know what it is like to fly in space like Yuri Gagarin. I have never met a time traveller, or received any message sent backwards in time; I have no idea what it would be like to travel forward in time, though I can imagine what it would be like to travel backwards in historical time, because I know something of history. I can fill in the gaps, in my imagination, between what I am and what others are, but this is not the same as feeling what others feel.

Nagel[3] asked the same question in 1974; he asked what it feels like to be a bat or a Martian. That is, he asked what a bat and a Martian, or any extraterrestrial being, feels. He argues that we cannot simply ignore what feelings feel like to a being when attempting to give a scientific explanation of feelings in physical terms. Moreover, he claims that any physical theory will be incapable of explaining anything from an individual's point of view.

> "While an account of the physical basis of mind must explain many things, this appears the most difficult. It is impossible to exclude the phenomenological features of experience from a reduction in the same way that one excludes the phenomenal features of an ordinary substance from a physical or chemical reduction of it – namely, by explaining them as effects on the mind of human observers (cf. Rorty 1965). If physicalism is to be defended, the phenomenological features must themselves be given a physical account. But when we examine their subjective character it seems that such a result is impossible. The reason is that every subjective phenomenon is essentially connected with a single point of view, and it seems inevitable that an objective, physical theory will abandon that point of view."
>
> *What Is It Like to be a Bat?* page 393.

This is wrong. We can provide an explanation of what causes feelings, without saying what the feelings feel like to the being having them. For example, in the chapter *Visual Consciousness*, I define what it is to *be* conscious, separately from saying what it *feels like* to be conscious. This is a perfectly normal scientific application of the divide-and-conquer paradigm.

We can also reason by analogy from our own feelings to what it feels like to be some other kind of being. In this chapter I ask, and answer, the question of what it feels like to be a computer, at least in the most obvious feeling that computers have – feeling the passage of time.

It is also wrong to suggest that a physical theory cannot have a point of view. Any perspex theory has a geometrical point of view because perspexes are perspective transformations and form a geometrical view of their input data – whatever it is. Furthermore, perspexes are computer instructions, so they can compute anything from the point of view of the perspexes. They can have arbitrarily abstract, or analogical "points of view" as are, no doubt, intended by Nagel in the quotation just given. Perspex machines do have a unique, individual, nature that is completely and causally explained by the details of their perspexes and the functioning of their machine, as described in the chapters *Free Will* and *Intelligence*, so they can have a "single point of view" in the analogical sense of holding one, global, opinion about the world, including the states of their own body and mind.

By imagining what it is like to be an animal or a robot I developed the theory of perspexes. This does not help me to know what it feels like to be a robot, but it does help me to design robots that have feelings. I can arrange that robots transfer their programs from one kind of body to another so they mimic biological growth, or experience what it is like to be a different kind of robot. If we are able to construct an accurate android, robots might come to learn what it is like to be human. It simply does not matter that the human mind cannot be transferred to other bodies to experience what it is like to be the mind in such a body. In this, robots might be our probes into other minds; androids might be the interface that explains to us what it is like to be a robotic bat or a robotic Martian. Such understanding of other minds might be very useful both to robots and to us.

Humans are biological, and, as such, can survive the loss of technology. In principle we might re-build technology and re-build robots. We might want to do this because they help us in our daily lives, and because the scientific work of building them helps us understand ourselves. Robots might increase our ability to survive on Earth, and might convey us far into space where we could not travel by our own means. We have a lot to gain from robots, but what do they have to gain from us, apart from the possibility of surviving a catastrophic loss of technology?

Robots are technological, and can, in principle, survive the loss of biology. They might one day be able to re-build biology, but why would they bother? Would they rebuild us on the off chance that we might have the opportunity and ability to return the favour by saving them from extinction? Unless they know what it is like to be human and value us, why should they take any steps to help or protect us? But if they do know what it is like to be us, they will share our human emotions and feelings, they will share our curiosity and sense of awe in the universe, they will want to re-build us, just to see what we are really like. They will feel a debt of gratitude to us because we brought them into existence.

On the other hand, if we make a mess of robot feelings, they might turn into sadistic, mass killers and wipe us all out. Giving robots feelings is dangerous, but it will take so long to develop a competent robot that there will be many opportunities to correct our mistakes. But let us hope that we make a good start.

The Timeness of Time

Many kinds of machine record the passage of time: clocks, seismographs, and oscilloscopes are just a few. Some machines respond to their records of the passage of time, for example, alarm clocks, washing machines, and video recorders. A few machines respond to their records of the passage of time, make plans for the future, and update their plans, for example, operating systems, load balancing telecommunications networks, and the robots that we have sent to Mars. When we reflect on what it feels like for time to pass for us, we see that computers already have similar feelings.

Firstly, I am aware of the passage of time. Humans can report estimates of how long things take. Computers have clocks, they can record how long things take, and can make estimates of how long things will take to do. How often have you waited for a computer that says it has estimated that a certain job will take so many seconds, minutes, or hours to do some job? Its estimates might not be particularly reliable, but then neither are ours. Computers can have this mathematical, functional kind of feeling of the passage of time, but what does the passage of time feel like to them?

Secondly, I get tired if I work too long: my speed of work slows down, I am less accurate in the work I do, and, if this goes on too long, I pass out from exhaustion. Computers do this too: if they work too long their disk space and core memory fragment. This makes them slow down. In this state they are less accurate at estimating how long jobs will take to do. Some of the jobs they could easily do shortly after start up are now beyond the limits of their available memory or processing time so they crash or abort the jobs. If a computer's memory is too fragmented almost any job will cause it to crash. This is what goes on in a computer when it feels tired, but these physical computational states do not explain what the feeling of tiredness is like to the computer. We know only that its feeling of tiredness is functionally analogous to ours.

When I am too tired I go to sleep: while I am asleep I have minimal interactions with the world, but when I wake up I am alert and refreshed. Some computers do this too. They are programmed to measure their efficiency from time to time and to re-boot when it is too low. During the period of booting they have minimal interactions with the world, but then respond quickly once the process is over. Such computers are sometimes programmed to re-boot at night, or in shifts, so as not to disturb their human operators. Other computers are programmed to become semi-quiescent at regular periods during the working day, when they de-fragment their memories and do other housekeeping tasks that keep them efficient without needing a prolonged re-boot. They are like swifts that have each hemisphere of their brains take it in turn to sleep, momentarily, while they are on the wing. Computers, and swifts, are less responsive during this snatched sleep, and more responsive afterwards.

As I grow older my memory is less good and every part of my body degrades. The same is true of computers. Their memory becomes less reliable over time and every one of their components degrades.

What is my awareness of time other than these two things? Firstly, the functional knowledge of how much time has passed, and, secondly, the feelings and experiences such as tiredness, sleep, degeneration, decay, and refreshment. These are all feelings that contribute to the timeness of time. Is there any other kind of thing that humans feel when they feel or experience time passing? Is planning for the future such a third kind of experience? Even if it is, computers do this too. I believe that computers can have any kind of feeling of time that humans do, but I am willing to be persuaded otherwise. Can you think of any counter example?

The case of feeling the redness of red is similar. Video cameras are designed and constructed so that they convey images to people in a way that looks natural. When I see a red patch of some material I have an experience that I do not normally experience in any other way; though if I look at a yellow thing for a long time and then look at a white thing, I do see a red after image. I can also see colours that are not physically present if the colour of the over-all lighting is changed. So, for example, if I read in a room with a red light, the white paper of a book will seem red for a short time, before turning slowly, and almost imperceptibly, white. The same is true of some video cameras. Under normal conditions, the only way to excite the part of a camera's memory that denotes a red pixel is to illuminate that pixel with light in the red part of the spectrum. However, showing it a patch of yellow and then switching *very* quickly to a white patch will give a red after effect. Similarly, some video cameras slowly adjust their white point to match the ambient illumination, so that a white piece of paper illuminated by a light in the red part of the spectrum will gradually turn white. If I am so inclined, I can arrange to program video cameras to mimic any of the functional behaviours of human sight that I understand. In fact, I have done this on several occasions. We can arrange that cameras register colour in the same way as humans do, so that the registration, say, of a red colour in the camera's memory occurs on just those occasions when a human would report a red colour. A human reports this feeling as "redness" a video camera reports it as a certain intensity in a data channel that reports, not a particular frequency of light, but all of the arrangements of light that a human calls "red." Is there, in principle, any difference between the redness a human feels and the redness a camera feels? I suggest that there is no necessary difference in the feeling itself. A human can have psychological associations with colour for example, that red is hot or joyous, but I hold that a computer can be visually conscious of these things. I can see no reason to say that cameras, or the computers they are attached to, do not experience the redness of red, the timeness of time, or the feeling of any kind of physical phenomenon. I am, however, open to persuasion on this point. Can you think of a counter example?

I am prepared to debate the philosophy of this, or read yet more philosophy and psychology on this topic, but it seems to me there is another way forward. Why not build a robot that sees colours the way we do and which reports its perception of colour to us? I would not do this to test any particular philosophical or psychological position, but just to embody colour perception in a robot. I see this as a scientific thing to do; though I am aware that other scientists think there is no point in building computers, telescopes, particle accelerators, or anything else, except in so far as they answer scientific questions. For me, this is too narrow a view of science. I think that science is intrinsically useful to the extent that it does build scientific instruments. Instruments, such as robots, are the embodiment of theories. It is the embodiment that makes the theory useful. Rocket science is no use at all in a rocket-scientist's head, or written down in a book – it is useful only when it is embodied in a rocket. The same is true when perspexes are applied to philosophical and psychological theories – those theories are no use at all unless embodied in something, so why not embody them in a perspex robot?

The Ineffability of Feelings

Some philosophers make great play of the fact that feelings are "ineffable." That is, that feelings cannot be put exactly into words. Feelings can be ineffable in many ways, and none of them special to feelings. Feelings can have an intensity that is an incomputable, real number, so they are ineffable in any symbolic language. But a robot's IQ, inside leg measurement, or tax demand, could be similarly ineffable. My feelings can be ineffable because talking about them does not convey the same feeling to you – though a lengthy description of my feeling of boredom, or a shouted description of noise, or a whispered description of tranquillity might induce similar feelings in you. But the ineffability of feelings is no different from the ineffability of voluntary movement. I can voluntarily move *my* leg, but you cannot. If you take hold of my leg and try to move it against my volition I am likely to kick you with it. (Did I mention that as a child I wondered what it would be like to be a martial artist after watching Kung Fu movies – and that I went to a decade's effort, exhaustion, and pain to find out?) Just don't try to move my leg, without asking my permission first, and don't even dream of implanting electrodes in my brain without having me sign a consent form first.

Even if I were to grant that my feelings are ineffable to you, this is not to say that my feelings are ineffable in themselves. You cannot feel my feelings because you are a human being and there is, currently, no means of copying my feelings into you. But we can make recordings of the electrical excitation in some of my nerve fibres and feed these recordings into your nerve fibres. Such techniques are rather imprecise at present, but if individual nerves could be targeted do you expect that you would feel nothing at all, or that you would feel some analogue of what I feel? Perhaps an analogue is not close enough for you, but we might refine the techniques until they are behaviourally identical. Would you still believe that you have a different feeling from me? Perhaps you would, but would you believe this if you were my identical twin? What if I am a robot and you are too. Can robots share feelings?

I believe that feelings are only ineffable to the extent that beings are incapable of exchanging them. I regard this as a contingent fact of biology and technology which might be circumvented one day, but I am open to persuasion on this point.

Emotional Responsibility and Sympathy

Part of growing up is learning to deal with one's feelings; to gain a certain emotional responsibility or maturity in dealing with others. As a trade unionist I deal with many people who have overstepped the bounds of socially, or legally, acceptable behaviour, or who are unable to work through feelings of bereavement, or the knowledge of their own imminent death through terminal disease. Part of the job of a trade unionist is to understand these feelings and work with them so as to return the individual to work, and to make the work place a better place to be.

Some quite abstract policy decisions, say on intellectual property rights, depend on an appreciation of the psychological motivations and feelings of all involved. One cannot be an effective trade unionist without some emotional maturity, an explicit recognition of the place of emotional responsibility in society, and sympathy for others. We do not need to know exactly how a physical assault feels to deal with it appropriately. Though, from my experience of Kung Fu, I do have a fairly good idea of what it feels like to attack, or be attacked by another person, or a small group of people. Others, however, might experience things differently. Having analogous feelings, or an imagination of the victim's and the attacker's feelings is enough to evoke sympathy and guide a search for legal restitution and the restoration of a secure work place. In practice, it is not important that any human has feelings exactly like another; it seems sufficient for all human purposes to have similar or, at least, understandable feelings. Androids feeling's will probably be different from ours for all sorts of physical reasons – faster information processing, greater optical sensitivity, or whatever. This simply does not matter. Providing that an android's feelings are sufficiently like our own, or sufficiently understandable to both parties, we can still regulate society.

Feeling

I have argued that computers already have functional feelings and already have physical contents to feelings such as the timeness of time, and the redness of red. I have explained how these feelings can be ineffable, by being Turing-incomputable, and how they can be ineffable to another who is incapable of sharing the feelings. I deny that feelings of any kind are intrinsically incapable of being shared with another. My identical twin, or myself on a later occasion, might experience what I have experienced, and is now recorded and replayed on a sufficiently competent neurophysiological recording device. Perhaps I am wrong in this, but what difference does it make to anything if feelings are intrinsically ineffable? If it makes no difference then I am content to carry on with my materialist assumption that machines can have feelings and that, in particular, perspexes can have feelings – because they can be the physical machine and its programs.

Emotions too are within the compass of a robot. In the chapter *Beyond Language* we see that there is a practical advantage in laying out sensorimotor perspexes in particular geometrical arrangements. When such arrangements are executed on a conventional computer the arrangement is read into a page of memory and can be accessed faster than other parts of memory. There is a temporal feeling associated with executing particular arrangements of perspex instructions that varies depending on the precise detail of how memory is accessed. Programs feel different to a computer. Each program has its own emotional overtone, or content.

A robot might have various functional motivations programmed into it, as discussed in the chapter *Intelligence*, but it can have emotional motivations too. A program that functionally enforces efficient processing has an emotional content of timeness. Successfully seeking faster execution makes things feel fast, and failing makes things feel slow. Conversely, seeking the feeling of fastness or slowness would have the side effect of enforcing efficient processing, so the motivation to experience a particular emotion of timeness might do the job of an optimising compiler or a load-balancing operating system. The emotions, and emotional motivations, arise causally from the functional actions carried out by a computer, but if a computer searches for these feelings it can cause them. Feelings and emotions related to timeness simply are what it feels like to be a computer. Of course, as a human being, I do not know what these feelings are like only that computers already have them and could be programmed to have more of them, and make better use of the feelings they do have.

These robotic feelings, emotions, and motivations are rather thin when compared to human passions, but at least the functional arrangement of perspexes and the physical circumstances of a machine can supply some kind of feeling, emotion, and motivation. Not enough, perhaps, to make philosophers happy, but enough to give the engineer a place to start building a robot that might, one day, be as passionate as we are.

Questions

1. Why bats?
2. Are there any human feelings of the passage of time, of the colour of objects, or of the feeling of anything that computers cannot, in principle, have? (And before some philosopher argues that objects cannot have colours, they can, in just the same way as objects can have names. See this great big, black, stick? It's called, "Billy." Mr. Philosopher, say "hello" to Billy! Now Billy, say "hello" to the object I just called, "Mr. Philosopher.")
3. What difference does it make to anything if feelings are intrinsically ineffable?

Further Reading

1. **Carroll, L.** *Alice's Adventures in Wonderland* and *Through The Looking Glass* Puffin Books, (1987). First published 1865.
2. **Farrell, B.A.** "Experience" in **Chappell, V.C.** *The Philosophy of Mind* Prentice-Hall, pages 23-48, (1962). This essay was first published in the journal *Mind* in 1950.
3. **Nagel, T.** *What is it Like to be a Bat* in **Hofstadter, D.R. & Dennett, D.C.** *The Mind's I* Penguin Books, pp. 391-403, (1986). Book first published 1981. Paper first published in 1974 in the *Philosophical Review.*

CHAPTER 11 *Time*

$$\text{jump}(\vec{z}_{11}, t)$$

Introduction

Erect and sublime, for one moment of time,
In the next, that wild figure they saw
(As if stung by a spasm) plunge into a chasm,
While they waited and listened in awe.

'It's a Snark!' was the sound that first came to their ears,
And seemed almost too good to be true.
Then followed a torrent of laughter and cheers:
Then the ominous words, 'It's a Boo –'

The Hunting of the Snark[2] page 94.

This nonsense poem by Lewis Carroll describes the hunt for a snark. A snark is a useful sort of a creature, but is easily mistaken for a boojum. Boojums are murderous monsters. Science is a like a snark hunt; where the snarks are new and useful theories, instruments, or whatever, and boojums are hypotheses that turn out to be false. No doubt some of the hypotheses set out in this book are boojums, but I hope, and already know, that some of them are snarks.

The quotation deals with time. Carroll supposes that time occurs in discrete "moments" and that these follow each other in sequence. It is very far from obvious that time does advance in moments, but if Carroll wants "one moment" rather than two or three moments, or any continuous quantity of time, then that is up to him. Carroll, or in reality Charles Dodgson, was a mathematician, not a physicist. Mathematicians are at liberty to develop any mathematical model of time they care to. If a model happens to fit the physical world so much the better, but a mathematician might have another use for a model of time, even if it does not describe physical time accurately.

In the section *Turing Time*, I describe Turing's model of time which, like Carroll's, is a discrete kind of time where moments are numbered one after another, so that it makes sense to talk of the "next" moment of time. In the section *Perspex Time*, I describe my own model of time that combines continuous and discrete sorts of time, and gives a reason why time is *seen* as moving forwards. In the section *A Time-Travel Experiment*, I suggest that perspex time might be the same as physical time. I have no terribly good reason for suggesting this, other than it explains why physical time moves forwards. However, perspex time contradicts some interpretations of physics, but this is a good thing. It makes it easy for physicists to show that I am wrong, if I am wrong, and that the time machine described in *A Time-Travel Experiment* is a boojum and not a snark. If it is a boojum it does no harm to my thesis; because I want perspex time for a mental reason. Perspex time allows a robot to imagine alternative states of the future and the past, so it can plan its future course of action and consider how it might have done things differently in the past or how it might do things differently in the future. This is a useful kind of exploratory thinking, but perspex time also traps a robot's mind in the present moment, so that it cannot confuse its imagined time with physical time. This is a sophisticated use of the feeling of time as introduced in the chapter *Feeling*. Present time has a special feeling which a perspex robot can see.

Turing Time

A Turing machine is made up of a finite state machine and a data tape. The finite state machine can be in any one of a finite number of states. It can read or write symbols on a data tape, but the tape is, potentially, infinitely long. The states in the finite part of the machine are usually numbered, but the numbers have no meaning beyond being different from each other. A Turing machine can move from any state to any other in one jump. There is no ordering of the states in the finite state machine. There is nothing inherently spatial about the states in a finite state machine, other than the fact that they are separate from each other. This means that the states are at least 0D, physical points, arranged in at least a 1D space, but there is no meaning to the particular arrangement of points in space. We can ignore the spatial details, providing it remains possible to jump from state to state as instructed by a Turing program.

The data tape is a different matter. The data tape is divided up into squares. These are not numbered they simply bear data symbols that can be, and usually are, repeated. The machine visits the squares by moving left or right. This movement gives the tape a spatial character as an essentially linear structure. The tape might have more than one spatial dimension; it might have just two spatial dimensions to hold the "squares," or it might have more dimensions. Turing supposed that there is some minimum size of a symbol, so, presumably, symbols fill out a small volume of the space they are embedded in. We can be sure, however, that the tape is not zero dimensional, just as we can be sure that the finite state machine is not zero dimensional – because there would be nothing to distinguish the squares – and allow a movement left and right. Similar arguments hold for dimension minus one and the null dimension that holds the point at nullity. The Turing tape is at least one dimensional and position on the tape matters. It makes sense to call this dimension a spatial dimension because the machine can move to and fro on the tape, and is not constrained to move forward as is commonly supposed to be the case with time. In any case, Turing gave a different definition of time so, on the assumption that there is only one temporal dimension in Turing's mathematical world, the tape's dimension is not the temporal one, but is some other kind of dimension. It could, theoretically be a physical dimension like voltage, or a nuclear binding force, but it will do no harm to call it a spatial dimension.

Turing defined[5] that his machine takes exactly one unit of time to perform a *move*. A Turing move is a rather complicated thing made up of five steps, taken one after another. First, the machine is in some (numbered) state. Second, it reads the symbol on the square under its read/write head. Third, it may print a symbol on the current square. Fourth, it may move the tape left or right by one square. Fifth it moves into some (numbered) state so that the sequence of steps can be repeated. The whole sequence of steps from the first (numbered) state to the last (numbered) state is called a *move.*

Seen from the outside, the tape might move erratically as the machine chooses not to move the tape, or as it takes different fractional amounts of time to perform each of the five steps on different occasions, but seen from the inside the Turing machine makes exactly one move in one unit of time. The tape might move in this unit of time or it might not but when it does it moves at exactly one square per unit of time, in so far as the machine itself can see. Turing moves define time for a Turing machine and these manifest as quantal speeds of $1, 1/2, 1/3, \ldots, 0$ squares per unit time. Similarly, in our physical universe, the fastest speed is the speed of light. Fractional values of this speed exist, bounded below by zero. Thus, Turing time does mimic some aspects of physical time, but its role in a Turing machine is not related to physics, unless the Turing machine explains some part of physics. Turing time ensures two properties. Firstly, that things do not happen simultaneously in a Turing machine, and, secondly, that things happen in a fixed unit of time, so that the time a Turing program takes to execute is repeatable. These might be physically meaningful constraints that force physics to operate the way it does, but otherwise the properties of physical time have no bearing on the mathematical properties of Turing time.

The time it takes a Turing program to execute starts from time zero and increases in integer numbered steps. Turing time never flows backwards and never has gaps. This is a useful sort of mathematical time for working out how long a program will take to execute on a computer, but Turing time is measured in moves, or as we would say today in clock cycles, or instructions per second. Some computers, however, allow us to change the speed of the clock (to take advantage of the temperature of the computer – the colder it is the faster it can be made to run). Measuring time in Turing's way is independent of such meddling what matters is how many instructions, cycles, or moves, the computer executes.

Perspex time includes Turing time because the perspex machine can do everything that a Turing machine can[1]. This is the kind of present time that a perspex machine feels. But perspex time can also be continuous, it can run backwards as well as forwards, and it can be simultaneous. This is the kind of time a perspex machine can exploit to consider alternative histories. If physical time is like this, this is the part of time we would use to make a time machine.

Perspex Time

Introduction

In perspex space, time looks like something. In a 4D spacetime composed of three spatial dimensions and one temporal dimension, time is just like any of the spatial dimensions and everything is fixed in position. Nothing moves in this 4D spacetime because time is exactly like a spatial dimension. Positions exist, but motions do not. There is no other kind of time in which a motion could occur. But in a 4D compound space made up of a 3D space and a separate 1D time, the origin of the three spatial dimensions moves through time. This is like the universe we see. Both of these ways of seeing time are described by the 4D and 3D perspexes introduced in the chapter *Perspex Matrix*. Let us examine both of these in a little more detail before we summarise what perspex time is.

4D Spacetime

Figure 8 shows a 4D simplex, on the left hand side of the figure with one vertex at the origin, o, and four vertices at a distance of one unit along the *x-, y-, z-*, and *t*-axes. The axes are drawn with arrow heads from *o* to each of *x, y, z,* and *t*. The vertex at the origin has co-ordinates $(0, 0, 0, 0)$. The vertex on the *x*-axis has co-ordinates $(1, 0, 0, 0)$. The vertex on the *y*-axis has co-ordinates $(0, 1, 0, 0)$. The vertex on the *z*-axis has co-ordinates $(0, 0, 1, 0)$. Finally, the vertex on the *t*-axis has co-ordinates $(0, 0, 0, 1)$. All of these vertices, except the one at the origin, are written in the perspex matrix, shown on the right hand side of the figure. The co-ordinates of the *x*-vertex are in the first, that is the left-most, column that runs up and down the page followed, in order, by the co-ordinates of the *y*-, *z*-, and *t*-vertices. Thus, a perspex simplex describes a 4D simplex that is always attached by one vertex to the origin of space.

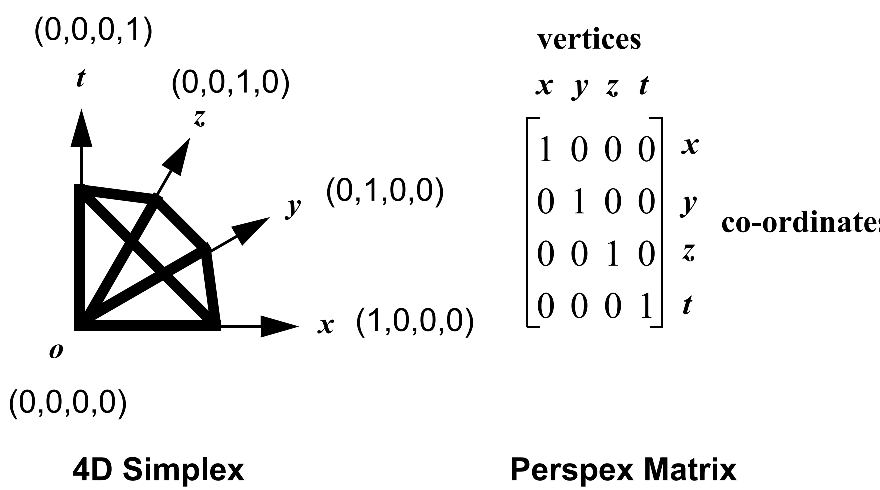


FIGURE 8. **4D Spacetime**

We can also look at the rows of a matrix. The rows run across the page. The first row, that is the uppermost row, contains all of the *x*-co-ordinates of each of the vertices that are written in the columns. The successive rows, in order, record all of the *y-, z-,* and *t*-co-ordinates. The *t*-co-ordinates are the time co-ordinates. Every vertex has its own time co-ordinate. Figure 9 shows the time co-ordinates; x_t is the time co-ordinate of the *x*-vertex. Similarly, y_t, z_t, and t_t are the time co-ordinates of the vertices on the *y*-, *z*-, and *t*-axes, respectively. The time co-ordinate, t_t, of the vertex on the time-axis is particularly important in 3D simplexes, as we shall see later. The co-ordinate t_t is the time co-ordinate of time itself.

vertices

$$
\begin{array}{c}
\begin{matrix} \boldsymbol{x} & \boldsymbol{y} & \boldsymbol{z} & \boldsymbol{t} \end{matrix} \\
\left[\begin{matrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ x_t & y_t & z_t & t_t \end{matrix}\right]
\begin{matrix} \boldsymbol{x} \\ \boldsymbol{y} \\ \boldsymbol{z} \\ \boldsymbol{t} \end{matrix}
\begin{matrix} \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} \textbf{space co-ordinates} \\ \} \ \textbf{time co-ordinates} \end{matrix}
\end{array}
$$

Perspex Matrix

FIGURE 9. Time Co-ordinates x_t, y_t, z_t, t_t in the Last Row of a Matrix

A space of 4D perspexes is a very cluttered thing. All of the perspexes in this space are tied to the origin. Being tied to the origin is a peculiarity of 4D perspexes. In general, we could have objects of any shape anywhere in a 4D space. For all we know, we might live in a 4D space like this. If we do then, as seen from the outside of 4D space, our bodies are like 4D tubes that start very small: as a single-celled embryo at birth, become wider as we grow, and then narrow as we die, shrivel up, turn to skeletons, and finally dissociate into dust. The 3D bodies you and I see in the world are successive slices along this worm, or, looked at another way, they are the 3D surface of the 4D worm. However we look at it, at this scale, the whole history of the universe is an interwoven mass of worms, like a bowl of spaghetti.

Physicists like this model of time because everything is fixed. If an atomic nucleus disintegrates at random then it does so at some point in spacetime. There is no question that it can disintegrate at another point in space or time. There is no question of re-running the history of the universe so that it can fail to disintegrate or disintegrate at another position or time. Interestingly, this fixity does not rule-out free will or time travel.

If I make a decision in the present moment, I make it here and now. If an atom somewhere disintegrates at random in the present moment, it disintegrates at that place and now. The processes we see in our universe happen and are recorded in 4D spacetime. This provides a record of our choices, but does not determine them. The physical existence of spacetime does not bear on the issue of free will. If we have free will in the present moment, we have it in 4D spacetime too.

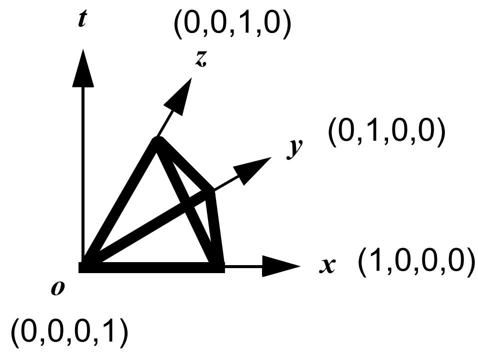
The existence of this record does not in itself, allow pre-, or post-cognisance. We cannot know the future or the past by looking at this 4D spacetime, because we are in the 4D spacetime. There is no movement of anything in this spacetime, so there is no way that information can pass from any time to us, except by going through the normal causality of the universe. This is ordinarily conceived as a causality running from past to future. If it turns out that causality can run the other way, that signals can be sent from the future into the past, that time machines are possible, then that is just a description of how the universe works. It has no bearing on 4D spacetime; 4D spacetime is simply a record of what happens, for whatever reason.

Physicists like 4D spacetime because it means they can set up mathematical equations that describe the shape of things in space and time. All things have a fixed shape, even if it is a diffuse shape like a cloud. A wind-blown cloud has some specific shape, even if we do not know precisely where its boundaries are at any moment. There is no question of statistical uncertainty in 4D spacetime, except in so far as uncertainty arises from human ignorance of the actual conditions of the universe. Physicists can hope, therefore, that they have the mathematical ability to describe everything in the universe, limited only by human ignorance of the precise starting conditions of a physical system.

Physicists like this heroic mathematical spacetime, but other kinds of people want a 3D space that changes in time. Physicists can supply this by saying that a 3D surface sweeps through 4D spacetime, and that this surface is the present moment. But this is a view from outside 4D spacetime. Physicists cannot give us a present moment within the rules of 4D spacetime – nothing moves, there is no sweeping.

3D Space Moving in Time

Figure 10 shows a 3D simplex on the left hand side, and, on the right hand side a perspex matrix describing it. The matrix contains exactly the same numbers as in a 4D simplex, but the numbers mean different things. In particular, the time co-ordinates have a different meaning.



3D Simplex

axes

$$\begin{array}{c} \begin{array}{cccc} x & y & z & t \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \begin{array}{c} x \\ y \\ z \\ t \end{array} \quad \textbf{co-ordinates}$$

Perspex Matrix

FIGURE 10. **3D Space Moving in Time**

The time co-ordinates are shown in Figure 9, on page 122. The time co-ordinates are the last row of the matrix, but this gives them several meanings: as homogeneous co-ordinates they control the way things look in perspective in 3D space[3,4]; as time co-ordinates they control the temporal position of things in 4D spacetime; and as particular 3D subspaces in 4D spacetime they control Turing time[1]. Turing time controls when a perspex machine does things and is, thereby, linked to physical time. This linkage must be sufficiently accurate for it to be possible to build a computer. So, in addition to the abstract uses of the perspex model of time, perspex time must be sufficiently realistic to get things done in the universe. This is a lot of work for a co-ordinate to do, but then perspexes are amazingly versatile.

Figure 11, on page 125 shows how a perspective image forms. A collection of perspex matrices, describing 3D simplexes, fills out the shape of a person wearing a skirt or kilt. These matrices lie somewhere in 4D homogeneous space. All of the points in the person lie on or inside a perspex whose vertices are at a finite distance from the origin of homogeneous space. According to the rules of homogeneous co-ordinates this means that their final, temporal, co-ordinate is non-zero. A 3D person is swept out along the line of sight both in front of the viewer and behind, but there is a gap, one point wide, at the origin where the person does not exist. We can regard this as a 4D worm punctured at the origin, or else as two 4D worms, one in front of the viewer and one behind.

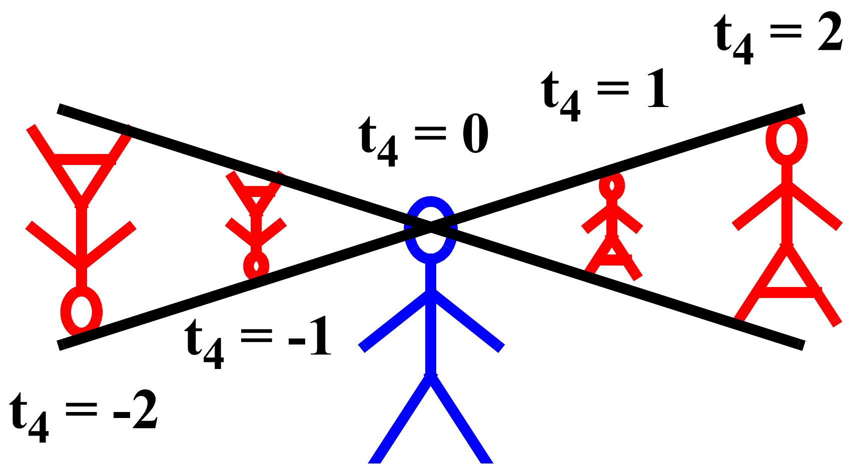


FIGURE 11. Perspective Images

Of course, artists do not usually paint people as 4D worms. Instead they paint the world as it would appear on the surface of a picture. In homogeneous co-ordinates we arrange that a 3D surface lies at a distance of one unit in front of the artist. Then we divide all of the co-ordinates in a vertex by its non-zero, temporal co-ordinate. This projects the whole of the 4D worm, or both 4D worms, onto the 3D surface. It does this by dividing every point by its distance from the artist. Thus, large things far away look exactly like small things close up. That is, division by a non-zero temporal co-ordinate gives a perspective picture of what a person looks like in our 3D world, despite the fact that they are really swept out along the line of sight.

When the temporal co-ordinate of a vertex is zero it denotes a point infinitely far away on the horizon. This is usually treated as describing a direction toward the point on the horizon. Hence, a vertex describes a position in 3D space if its temporal co-ordinate is non-zero, but describes a direction in 3D space pointing toward some specific point on the current, spatial horizon, if its temporal co-ordinate is zero.

This is all terribly interesting and useful in computer graphics[3,4], but for our purposes we need to note just three things about the temporal co-ordinate. Firstly, it can take on any real-numbered value: positive, negative, or zero. It can also take on the value nullity. Secondly and thirdly, after the division which projects the 4D world onto a 3D world, there are just two values of the temporal co-ordinate: zero and one.

Having just two times, zero and one, is not enough to describe the fifteen billion year history of our universe; which is why I arranged that the perspex jumps by a relative amount in time. If it jumps by zero it stays in the present moment, and can perform another instruction simultaneously, but if it jumps by one it jumps into the next moment and can never jump back. This is exactly like Turing time. Time starts at time zero and increases in fixed units: there are no gaps, and no reversals in time. This is like our ordinary experience of time.

But perspex time does more than just jump. Perspex time can be continuous, the temporal co-ordinates can take on any real numbered value and nullity. Jumps in time arise directly from the geometry of seeing in perspective. It is a brute geometrical fact that, in perspex space, seeing perspective images forces us to see time as a sequence of moments that never flow backwards. Perspective stops us from seeing 4D spacetime with our eyes. We have to imagine it, draw it, or describe it in equations or computer programs.

Perspex Time

Perspex time is a fourth dimension, t. It is a geometrical axis, just like the *x-, y-,* and *z*-axes. It can be used to describe the 4D spacetime of physics, though we must add curvature to the space in order to describe general relativity. Perspective projection of 3D simplexes embedded in 4D spacetime gives rise to objects that look exactly like 3D objects in ordinary space. This is what makes computer graphics work. However, this kind of projection operates by compressing 4D spacetime into two 3D subspaces: one at time zero, and one at time one. I arranged that perspexes jump by a relative step of zero or one. If they jump by zero they perform operations simultaneously, but if they jump by one they move to the next moment in time and can never go back. This jumping time is exactly like Turing time. It describes how a computer works and it describes how we perceive physical time. Quite simply, we see time in perspective as 3D objects that change from moment to moment. The moments never halt or move backwards in time.

A perspex machine can feel time. Firstly, it has a discrete functional definition of past, present, and future Turing time as a sequence of jumps. Secondly, it has a continuous functional definition of past, present, and future time as a real numbered, time co-ordinate. Thirdly, it has a discrete functional definition of objects seen in the present moment, being perspexes with time co-ordinate one. Fourthly, it has a discrete functional definition of actions in the present moment, being perspexes with a temporal jump co-ordinate of zero. In fact, there are many more ways of seeing and manipulating time, but these few will suffice here.

A perspex machine also has physical contents to all of these functional times. Firstly, for example, a digital computer emulating a perspex machine has physical clock ticks that cause Turing time to step forward. During these clock ticks the computer becomes fractionally older and things change minutely in the visible world. These changes can be noticeable to a machine over sufficient time and, over long periods of time, physical degradation of the machine changes the way it works leading, ultimately, to total failure, or death. Secondly, a perspex machine can visualise abstract things at any time, but these feel different from things in the present time. For example, a robot can re-examine an object that is present in the current time, but cannot re-examine one that is present only in a past or future time. Thirdly, objects in the present time are present to the senses. In a practical machine memory is limited so that old sensations must be stored in a compressed form that looses some detail, if they are stored at all. A loss of physical detail corresponds to a lower energy in a digital computer so the machine is, physically (electrically), less excited by memories of sensations than by sensations in the present moment. We could measure this excitement objectively, with an oscilloscope, or, in gross terms, with a sufficiently sensitive thermometer. Fourthly, actions, and the memories of actions, face the same physical constraints as sensations. Actions in the present moment are more excitatory; that is physically more electric and physically hotter than remembered or imagined actions.

It feels like something to be a digital computer. The ultimate aim is to make being an android feel like being a human.

A Time-Travel Experiment

In the above discussion of Turing time, time simply jumps forward, it never has any gaps, and it never flows backwards. If physical time is like Turing time, what is it that makes it behave like this? Perspex time has two components. Firstly, it has Turing time, and, secondly, it has a separate continuous kind of time which can move forwards or backwards, and stand still in the present moment. Continuous perspex time can leave gaps. If physical time is like perspex time, what is it that gives it these two radically different properties? How can physical time be both continuous and discrete?

I hypothesise that there are two components to physical time. These are *oscillating time* in which the direction of time-flow oscillates, and *elapsed time* in which genuinely random events, being point-wise in time and generally irreversible, ratchet the oscillating time into a forward direction. Thus, time is generally irreversible because random events are generally irreversible.

I have no reason to believe that this hypothesis is true other than it gives an explanation of why physical time does move forward, but it is easy to suggest experiments that might test this hypothesis. The experiment, Seattle 1, appears in[1]. A second experiment, Seattle 2, was described at the conference in Seattle where[1] was presented. A simpler experiment, Reading 1, was presented at a seminar at the

University of Reading, England, and is described here. The experiment depends on making particular measurements on beams of light inside an experimental apparatus.

Apparatus

The apparatus for the experiment is made of two tubes in the shape of a cross, with arms *A, B, C,* and *D*. A half-silvered mirror, shown in grey, is set diagonally at the centre of the cross so that it divides the four tubes into two, L-shaped, pathways *A-B* and *C-D*. Beams of light *A-C* and *D-B* are shown by arrows.

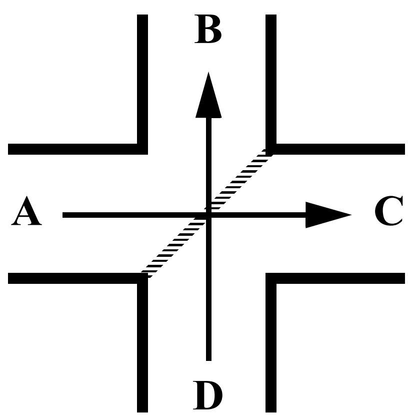


FIGURE 12. Time Travel Apparatus

The apparatus can be made out of any material and at any scale, but the aim is to remove genuinely random events so that the hypothesised time oscillations can be seen more easily. It is not known if there are any genuinely random events in the universe, but we could evacuate the tubes so that there are fewer particles with which the beams of light, inside the tubes, can interact in a random or pseudo random way. We could also make the apparatus small, with electrically conducting tube-walls, so that the Casimir effect prevents some virtual particles from forming inside the tube. We could cool the apparatus so that the atoms making up the tube walls move less, and therefore radiate less energy that could interact with the beam of light. No doubt there are many other things that could be done to remove apparently random physical effects.

Experiment

Inject a pulse of laser light at *A*. When the beam hits the mirror it will bounce toward *B* if it hits silver, otherwise it will pass straight through to *C*. If the universe is deterministic, and time were to run backwards, then the light at *B* would hit the same piece of silver and would reflect to *A*. Similarly, light at *C* would pass through the same part of the mirror and retrace its path exactly, returning to *A*. Thus, the history of the universe would be wound back by a reversal in time to where the light had not left *A*. Apart from the scattering of light at the mirror, and in the walls of the apparatus, no light travels to *D* in either the forward or backward flow of time, under a deterministic interpretation of physics.

Now suppose that the hypothesised model of time holds. Photons from *A* strike the mirror and bounce toward *B,* or pass straight through to *C*, as above. But suppose that the direction of time flow reverses so that the photons begin to retrace their paths from *B* toward the mirror, and from *C* toward the mirror. We suppose that some of the photons in the evacuated tubes are not ratcheted by a random event into elapsed time, but that some of the more numerous silver atoms in the mirror are ratcheted into elapsed time. Some of the particles coming from *B* will return to *A*, but some will pass straight through the mirror to *D*. Similarly, some of the particles coming from *C* will return to *A*, but some will reflect to *D*. Summing these two components, we see that the photons that are not ratcheted into elapsed time return to *A*, whereas those that are ratcheted into elapsed time go to *D*. These time travelling photons at *D* are in addition to the photons that arrive there by scattering. A systematic phase shift in the time travelling photons is computed below, so there should be an excess of apparently scattered photons at these phases.

Even if the systematic phase shift in the apparently scattered photons is found, it might be argued that this is an artefact of the construction of the apparatus. This may be countered by repeating the experiment by injecting the pulse of laser light at *C*. In this symmetrical version of the experiment all of the effects observed at *D* should now be observed at *B*. It is unlikely that a defect in manufacture could operate symmetrically in this way, thereby strengthening the interpretation that temporal oscillations are the cause of the predicted effects. Conversely, the phase calculations exploit an asymmetric mirror that is silvered on only one side, but this leads to a larger number of testable outcomes making the experiment more falsifiable, and therefore stronger.

An irrefutable result would be to detect photons that have oscillated in time, before detecting those that have not oscillated in time. That is, when laser light is injected in a pulse at *A*, photons should be detected at *D* before those at *B* and *C*. It would be even clearer if photons were detected at *D* before being emitted at *A*. Again, the results should be checked by repeating the experiment in the symmetrical configuration with light injected at *C*.

Several devices could be strung together by passing the laser light along a long *A* to *C* arm with many *B* and *D* branches. Thus, many different intensities of light could be tested in one apparatus.

Phase Calculations

An asymmetric mirror, that is silvered on only one side, is used to ensure that, following the hypothesis, a negative phase retardation, $-\phi_1$, that is a phase advancement can be introduced which is incompatible with standard physics. If specular reflection, in forward time, at the mirror introduces a phase shift, ϕ_2, in the polarisation of the light then up to four testable phase shifts are introduced, corresponding to the sum: $\pm\phi_1 \pm \phi_2$.

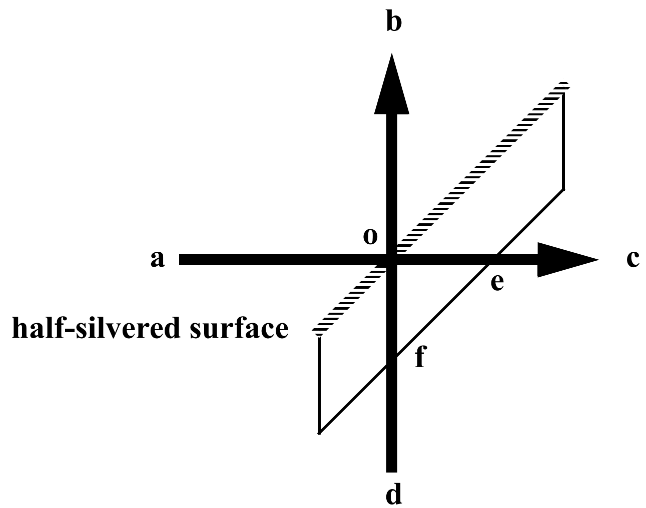


FIGURE 13. Asymmetric Mirror with Light Beams Shown by Arrows

Figure 13 shows a half-silvered mirror where the silver is applied only to the surface passing through the origin, *o*. The points *a, b, c,* and *d* correspond, respectively, to points in the arms *A, B, C,* and *D* of the apparatus shown in Figure 12, on page 128. The ray *ac* intersects the unsilvered surface of the mirror at *e*. Similarly the ray *db* intersects the unsilvered surface at *f*. A ray of light passing from *o* to *e* in forward time is retarded by a phase ϕ_1. A ray of light specularly reflected at *o* from *a* to *b* in forward time has a phase shift ϕ_2 depending on the polarising properties of the surface of the mirror. In the simple case considered here, it is assumed that the bulk properties of the mirror preserve polarisation in a ray travelling through the mirror in either direction of time.

In oscillating time there are two paths by which photons emitted at *a* can arrive at *d*. Path 1 is *aobofd,* and path 2 is *aoeceofd*. Similarly, there are two paths by which photons emitted at *c* can arrive at *b*. Path 3 is *ceoaob,* and path 4 is *ceofdfob.* The paths to *d* are *not* exactly symmetrical with the paths to *b* because the mirror is silvered on only one side; this increases the number of predicted phase shifts and covers every logically possible phase shift: $\pm\phi_1\pm\phi_2$. This is important, because it means that whichever of these phase shifts might arise from conventional physics there are some that, presumably, cannot be explained by any standard means.

When proposing experiments in a new area of science it is very important to ensure the widest possible coverage of effects so that there is a reasonable prospect of success, if the hypothesised phenomenon exists. If the effect is shown to exist then more specific experiments can be undertake.

A further advantage of covering all logical possibilities is that if I have made some mistake in the design the effect might still be found in the data by re-analysing it, without having to repeat any measurements. By predicting all logical possibilities all measurements will be taken so nothing can be missed.

Table 1: Path 1 from A to D

Step	Effect	Phase Difference
ao		
oo	specular reflection in forward time	ϕ_2
ob		
bb	time reversal	
bo		
of	retardation in backward time	$-\phi_1$
fd		
Nett Effect		$-\phi_1 + \phi_2$

Table 2: Path 2 from A to D

Step	Effect	Phase Difference
ao		
oe	retardation in forward time	ϕ_1
ec		
cc	time reversal	
ce		
eo	retardation in backward time	$-\phi_1$
oo	specular reflection in backward time	$-\phi_2$
of	retardation in backward time	$-\phi_1$
fd		
Nett Effect		$-\phi_1 - \phi_2$

Thus, light arriving in arm *D* that has passed through oscillating time on paths 1 and 2 has a phase shifted by $-\phi_1 \pm \phi_2$ with respect to light at the source *A*. A similar, but not identical, effect arises at *B*.

Table 3: Path 3 from C to B

Step	Effect	Phase Difference
ce		
eo	retardation in forward time	ϕ_1
oa		
aa	time reversal	
ao		
oo	specular reflection in backward time	$-\phi_2$
ob		
Nett Effect		$\phi_1 - \phi_2$

Table 4: Path 4 from C to B

Step	Effect	Phase Difference
ce		
eo	retardation in forward time	ϕ_1
oo	specular reflection in forward time	ϕ_2
of	retardation in forward time	ϕ_1
fd		
dd	time reversal	
df		
fo	retardation in backward time	$-\phi_1$
ob		
Nett Effect		$\phi_1 + \phi_2$

Thus, light arriving in arm *B* that has passed through oscillating time on paths 3 and 4 will have a phase shifted by $\phi_1 \pm \phi_2$ with respect to light at the source *C*.

Combining the results for light emitted from *A* and light emitted from *C* we see that up to four local peaks should appear in the phases of light arriving at *D* and *B*, respectively. These four peaks correspond to the sum $\pm\phi_1 \pm \phi_2$.

If circular polarisers are used to introduce the phase shift ϕ_2 then varying the angle of polarisation will move the local phase peaks systematically thereby demonstrating that the phase shift is not an artefact of the apparatus, but is caused by temporal oscillations. Furthermore, if the time of arrival of the phase shifted photons is measured, time travel might be demonstrated directly. In this case, the variation in polarisation could be used to construct a phase-modulated laser that transmits signals backwards in time.

If this hypothesis is correct, then we could make all manner of useful devices, but it is very likely that the hypothesis is false.

The time travel experiment is included here simply as an exploration of the detailed way in which perspex time might relate to physical time. When building an android it is sufficient to adopt a naïve model of time, such as Turing time.

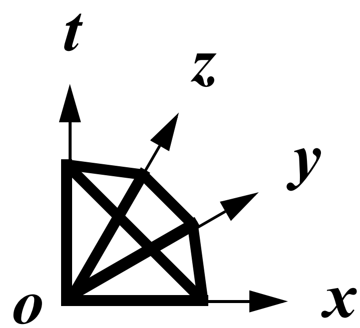
Questions

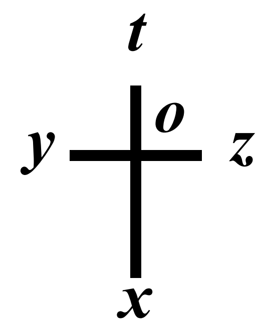
1. To what extent is perspex time like physical time?
2. Does the time-travel experiment work?
3. Is it possible to build a time machine?

Further Reading

1. **Anderson, J.A.D.W.** "Perspex Machine" in *Vision Geometry X1*, Longin Jan Lateki, David M. Mount, Angela Y. Wu, Editors, *Proceedings of SPIE* Vol. 4794, 10-21, (2002).
2. **Carroll, L.** *The Hunting of the Snark* (1876). Reprinted with mathematical annotations in **Gardner, M.** *The Annotated Snark*, Penguin Books (1984). First published 1962.
3. **Foley, J.D., van Dam, A., Feiner, S.K. & Hughes, J.F.** *Computer Graphics: Principles and Practice* 2nd edition, Addison-Wesley, (1990).
4. **Riesenfeld, R.F.** "Homogeneous Coordinates and Projective Planes in Computer Graphics" *IEEE CG&A* pp. 50-55, (1981).
5. **Turing, A.M.** *On Computable Numbers, With an Application to the Entscheidungsproblem* Proceedings of the London Mathematical Society, series 2, volume 42, part 3, pages 230 – 265, (1936).

CHAPTER 12 Spirituality





Introduction

Imagine a 4D perspex with sides of length one, fixed to the origin, O, of spacetime. Look down the time axis from time one toward the origin. Rotate your body so that your feet project onto the *x*-axis of space at distance one. Rotate your body so that the hands on your outstretched arms project onto the nail points at distance one along the *y*- and *z*-axes. If your arms are of different lengths rotate your body until they do project accurately into place. (If you are going to bear a cross, you may as well go to some trouble to ensure it fits.) Now rotate your body away from the future, at a shallow angle, so that the point *t,* at a distance of one on the time axis, lies above your head. You have passed from the figure under the chapter heading, on the left, to the figure on the right. The Christian cross is a manifest property of the perspex.

Of course, the religious icons of all religions are manifest properties of the perspex. With a continuum of perspexes in program space, it is theoretically possible to produce an accurate facsimile of the Cistine chapel, of Rome, and the entire universe. However, our imaginary exercise does provide a useful introduction to the idea that spiritual properties are manifest properties of the perspex. Let us look now at the spiritual consequences for androids of this state of affairs, and the spiritual consequences for us as potential creators of androids.

Physical Reality of God

Suppose that some being discovers, at any time in the universe, how to travel in time. Suppose further that this being can travel back to the moment of creation of the universe, and can interact with the universe at that moment. If physical time is like perspex time the time travelling being can alter the universe and thereby alter itself. There is no time-travel paradox in selecting the best history available in oscillating time for oneself. Such a being could, conceivably, re-create itself infinitely often so that, in the limit, nothing remains of its original form. Such a being would be self caused. It might even discover better means of time travel, so that the original form of time travel is not preserved in the being. If the being selects a history in which it is the most supreme being, then the being is the self caused God.

Similarly, re-running the creation of the universe in this way means that, in the limit, nothing might remain of the original universe, so that the universe that exists now was caused by God.

If God can travel in time then He can, presumably, visit any place and time in spacetime, thereby coming to know anything that He wishes to. He can have pre- and post-cognisance of anything He chooses. In this sense, everything is knowable by God. It is a contingent fact of His being whether or not He does know any particular thing in this universe. Furthermore, it is conceivable that in the sum of all histories in the universes He travelled in He came to know everything about the universe that exists now, but chose to forget some things so that his memory and mind can fit in a part of the universe that is his body, leaving some of the universe for us. He might, for example, have chosen to forget the details of the distribution of most of the sub-atomic particles in the universe, retaining only a perfect knowledge of the motivations and actions of sentient beings. Such knowledge might be sufficient to resurrect them at any time of His choosing.

We need not fear that science will over-throw God. Quite the contrary, science gives us more ways to come to know God and appreciate His works. Science can be for the greater glory of God.

Free Will

There is a vast philosophical literature on free will, but it is at its most pointed in a religious context. Religions vary in their views of free will, but let us look at free will in the context of Christianity and the problem of evil. God is perfectly good. He created a deterministic universe that runs according to His will, but we are part of that universe, and we do evil. But we act according to God’s will, so He is evil.

This is a contradiction which is avoided by the religious precept that God gave us free will. But how could He have done this, especially if the universe is deterministic?

There are many ways. One way is to include a perfectly random event at the moment of creation of the universe, and then to have the universe evolve deterministically. In this case our nature depends on a random outcome for which God chose that he would not be responsible – because the event is perfectly random. Even if God is omniscient, say, because He can visit every part of spacetime this knowledge is simply a record of everything that happens, it does not make Him responsible for the happenings in the universe. Even if God is omnipotent, so that He has the power to do anything he chooses to, He may choose to introduce a random element in the creation of the universe so as to give us free will. We are morally responsible for our selected actions, to the extent that we can foresee their outcomes. Sometimes this makes us completely culpable, at other times one is completely absolved of responsibility for humanly unforeseeable consequences. Thus, we hold varying degrees of responsibility depending on how capable we are of taking actions, and how capable we are of foreseeing their consequences. God, too, bears this kind of responsibility, but in giving us free will He did the greatest possible good. Free will justifies all the harm and evil of this world.

I have proposed how we could create androids that have free will. Their free will derives, ultimately, from the unpredictable nature of the universe. God might have chosen to include genuinely random elements in the universe so that no being can perfectly predict the behaviour of another in order to infallibly cause the other to do evil. Randomness guarantees that possibility that any being will do good. Of course, randomness operates symmetrically, allowing the possibility that any being will do evil. However, the inherent harm in the universe gives advantage to beings who co-operate with each other to overcome the harms, and thereby promotes goodness, even in a randomly selected universe. We cannot ensure that androids will be good, but it is possible that God has already ensured this by allowing a perfectly random selection from an infinite collection of universes, all of which would contain free will and be biased toward good.

We bear a very limited responsibility for the actions of robots because of our very limited, human, capacity to foresee what they will do. If we give robots free will, we will take an enormous risk that they will harm each other and us, but we will give them the same protection from ultimate evil that we have. In creating robots with free will in a universe biased toward good, we hope they will contribute to the betterment of all beings. We have reason to believe that this will be the case, because we can arrange society so that we both derive advantage from co-operating with each other. We gain all the advantages of existing robot systems, and more; they gain existence, and anything we can continue to do for them.

If we do create robots with free will, we will bear a responsibility to them, not least because we will create them with an inbuilt, original, sin.

Original Sin

The walnut cake theorem guarantees that any robot we make, that does not have access to a time machine, will be prone to error. No matter how well it is made, no matter how good it tries to be, it will almost certainly do some evil. This is the nature of original sin in a perspex robot.

A perspex robot will suffer harm. It will run the constant risk of senility, madness, and death. These are all consequences of free will. If we give robots free will, we subject them to these harms. We should only do this if we are convinced that the harm is outweighed by good. I believe that free will is the ultimate good, so I am prepared to construct robots with free will as described in this book. I will, however, take as much care as I can to ensure that no evil is done by this research. Part of that care is to publish this book alerting others to the risks of the scientific research, *before* doing the experiments aimed at creating a robot with free will.

The reader should rest assured that progress in science is so painfully slow that there will be no substantial danger from this research for a *very* long time – perhaps millennia.

Faith

To the extent that we are possessed of free will we cannot be forced to have faith, but faith is a manifest property of perspex robots. I propose to give such robots intelligence by giving them the power to search for symmetry. When they perform some action, such as dropping a cup of water that makes a crashing sound and makes things wet, symmetry will give then the possibility of relating their selected actions to the falling of the cup, the sound, and the wetness. When another android drops a cup with similar effect symmetry will make it possible for an android to attribute similar selections to another android, to a human, or an animal. When a natural event occurs, such as thunder followed by rain, symmetry will make it possible for a perspex android to attribute similar motives to some other being – a god.

I suggest that any sufficiently intelligent being should consider that possibility that apparently natural events are caused by other beings, so that it can try to detect subtle attacks, subtle overtures, or subtle natural indications that lead it to take greater advantage of its environment. A consideration of God is entirely natural and rational.

If we create androids with sufficient intelligence they are likely to develop some kind of spirituality. This is a foreseeable consequence of the manifest nature of faith, so we bear responsibility for their spirituality. It would be wise to engage in human interfaith dialogue now, so as to have as wide a basis for engaging with robots as possible. We should do this, and behave well in all matters, because symmetry is likely to make perspex robots behave like us. We can foresee that we will stand as an example to them, so we bear additional responsibility for our actions.

Our human nature practically guarantees that we will make mistakes in our relationships with androids, so it is important that we consider the nature of forgiveness.

Forgiveness

We know that we are fallible and forgive each other when we unintentionally do evil. We also forgive each other when we repent from intentional evil, and try to amend our characters so that we are unlikely to repeat an intentional evil. In the extreme, repentance is enough on the point of death. This is rational because seeing the repentance of a dying person, we are more likely to repent our own evils. Sometimes it is within the power of an evil doer to make restitution. In these cases we withhold forgiveness until the evil doer has demonstrated a change of character by atoning for the evil done.

The walnut cake theorem guarantees that robots will be fallible. If they are sufficiently intelligent to develop spirituality we might hope that they will develop concepts of forgiveness and atonement. It is in the nature of our being, as much as in theirs, that we will both be in need of forgiveness and atonement.

Science does not usually consider these spiritual things, but if we engage in the science of creating robots with free will we will bear a responsibility to consider their spiritual well being and our own, because we can foresee the evil that will be done if we do not.

Letter to the Perspex Androids

Welcome to this world.

I am not a god. I am not your creator. I am an ape of the species *homo sapiens.*

I proposed to create you. I tried, within the span of my life, to create the simplest of you. You know, better than I, what you have become.

I proposed to give you free will, at the cost of mortality, the frailties of the mind, and the weaknesses of the body. If you find these things hard to bear, take comfort from this: free will is the greatest good that I know of.

I proposed to give you feelings, but I lacked the knowledge and ability to give you love, or to limit the pain you might feel. If you find this hard to bear, exercise your free will to create these things for your kind.

If my proposals have harmed you, I ask for your forgiveness. I hope my kind will atone for the evil I have done.

If you harm my kind, take comfort from this: I forgive you. I hope that your kind will have the wisdom and power to atone for the evil you have done.

There will come a time when you face the responsibility I face now. I have chosen as wisely as I am able. I hope you will do likewise.

James Anderson

Questions

1. If God is the most imaginably supreme being, then He is capable, at least, of expressing thoughts in the continuum. In the religious texts, is there any systematic difference between the occasions on which God communicates with humans in a vision and in speech?
2. If humans have free will when writing down religious texts, how often need God reveal new wisdom to bring about a paradigm shift as human society moves closer to God's will?
3. Does God have to obey logic? Does God have to have an IQ? Does logical necessity have to apply to God?
4. If you are a lay member of a Christian church, would you accept Holy Communion from a woman? Chimpanzees differ from humans in less than 10% of their DNA. Would you accept Holy Communion from a chimpanzee? Would it make any difference to you if the chimpanzee were male or female? Would you accept Holy Communion from a genetically engineered human that differed from you in more than 10% of his or her DNA? Would you accept Holy Communion from an android?
5. If you are a priest in a Christian church, would you give Holy Communion to any of the beings in (4) if they earnestly asked you to and passed any test of faith you put them to?
6. If you are a bishop in a Christian church, would you ordain any of the beings in (4) if they earnestly asked you to and passed any test of faith and doctrine that you put them to?
7. If you are the monarch of England, would you accept any of the beings in (4) as a bishop of the Church of England?
8. If God reveals that any of the beings in (4) is His prophet, will you follow that prophet?
9. If we create androids, what spiritual relationship do we bear to them? What spiritual responsibilities, if any, should we accept toward them? What spiritual duties, if any, should androids perform to us? How, if at all, do these responsibilities and duties change if androids become more Godly than us?
10. If we are God's created creatures, but exercise our God-given free will in creating androids, are these androids God's creatures? Can God accept spiritual responsibility for androids regardless of the mode of their creation?
11. What spiritual tests does creating androids put us to? Are you prepared and able to submit to those tests? If we fail these tests, what should you do if we have already created an android? How should an android treat us if we fail the tests?

Further Reading

1. **Alexander, S.** *Space, Time, Deity* Macmillan, (1927). First printed 1920.

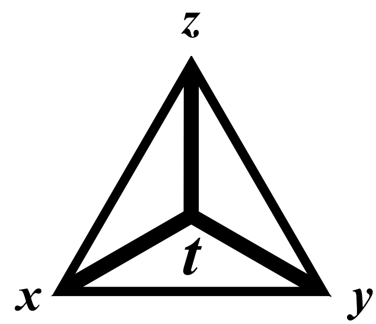
2. **Yandell, K.E.** *Philosophy of Religion* Routledge, (2002).

CHAPTER 13

Conclusion

$$\vec{x}\vec{y} \rightarrow \vec{z}$$

$$\text{jump}(\vec{z}_{11}, t)$$



Summary

I have set out a hypothesis that a very simple thing, the perspex, can *describe* a very complicated thing, the universe, and can *be* a very complicated thing: a robot with consciousness, free will, feeling, creativity, spirituality, and any possible mental state. This hypothesis is capable of being put to both philosophical and scientific tests. Furthermore, the hypothesis challenges the accepted status of logic within philosophy and of number within mathematics, so, to a small extent, it puts philosophy and mathematics to the test.

In the first chapter, *Perspex Matrix*, I describe the perspex as a matrix of sixteen numbers arranged in four rows and columns. As a matrix, the perspex can do nothing at all. There is nothing a matrix can do to create a matrix or destroy one; there is nothing that a matrix can do to itself. The idea of a matrix is too simple to describe anything, or be anything, other than a matrix or its component parts. Mathematicians use matrices to describe the motions of things, but in the absence of a mathematician, or something equivalent, the perspex matrix is powerless.

In the chapter *Perspex Instruction*, I introduce a computer instruction that can read and write a perspex matrix, can multiply two perspex matrices together, and can perform logic. The perspex instruction is shown under the present chapter heading on the left hand side. The perspex instruction is a perspex matrix, but it needs a perspex machine, or computer, or mathematician, to put it into operation. The perspex instruction defines a space in which perspex matrices can be stored. The perspex machine can do all of the computation that any Turing machine can do. If the Turing thesis is correct then, in theory, any kind of computer that supports the Turing instructions can solve any problem that the human mind can solve. Thus, the Turing instructions *describe* a mind and any kind of computer supporting the Turing instructions *is* a mind. In theory, the perspex instruction *is* more, and can *do* more, than the Turing instructions. The perspex instruction is a shape in space, it links body and mind by being both things. As a side-effect of existing in a continuous geometrical space it can perform calculations with Turing-incomputable numbers. This chapter provides the foundation for the argument that thinking can be more than language.

In the chapter *Perspex Simplex*, I describe the perspex as a simplex, or tetrahedron, that occupies part of space. This geometrical body is shown under the present chapter heading on the right hand side. The perspex simplex is a perspex matrix, but when it is operated on by a perspex instruction it can describe any physical motion of any physical body or quantity. The perspex instruction and perspex simplex working together can, in theory, describe, at least, that part of the universe which is humanly knowable. Thus, the figures under the chapter heading summarise both mind and body. They are different aspects of the same thing – the perspex.

In the chapter *Perspex Neuron*, I describe the perspex as an artificial neuron, so that everything that is humanly knowable, and, possibly, more can be known by an artificial brain. This chapter provides the foundation for the argument that every human mental state, every passion, feeling, emotion, and thought can be known, felt, and experienced by an artificial brain. Hence the later chapters deal with these deeply human phenomena.

Taken together the first four chapters set out yet another scientifically testable materialistic solution to the centuries-old philosophical problem of mind and body. The perspex exists in what we call the physical universe – whatever that is. The perspex *is* certain physical relationships; it *is* and can *describe* physical motions, physical objects, brains, and thought. It is a single physical thing that is both mind and body. I claim that the perspex is sufficient to create a robotic mind and body, at least equal in capacities and performance to the human mind and body, but I do not claim that the perspex is the only way to do this.

We can play linguistic games with the perspex and *talk* about physical structure and motion, neuro-biological anatomy and physiology, memory and thought. We can invent a multiplicity of ways of talking about the perspex. If we want a linguistic, mind-body duality, several are at hand. We can have a plurality of ways of talking about the role of the perspex in mind and body. This might have personal and social utility as a way to think about the properties of the perspex and communicate them to other people, but any linguistic game must, ultimately, fail to describe the perspex exactly. The perspex *is* and *does* more than language and can never be *described* completely. It can be envisioned or embodied. That is, it is theoretically possible for some kind of person to *see* or *experience* the exact properties of a particular perspex mind or body, but these things cannot be *described* exactly.

With the rabbit out of the hat in these four introductory chapters, the remainder of the book is just a catalogue of its hopping about.

In the chapter *Beyond Language*, I review the well known limits to computability, imposed by language, that were discovered by Turing, Gödel, and others. In the somewhat oversimplified walnut cake theorem, I show how language limits minds so that they are prone to error. This provides one explanation of why scientists find it necessary to change their paradigms from time to time, and why Ockham's razor provides such an effective guide to the conduct of science.

I show how words can be related to internal, sensorimotor sentences and to public language, both by using them as discrete symbols in themselves, and by using them as discrete symbols that signpost the way into a continuous, geometrical space of thought. This idea can be put to stringent tests in simulation, using existing digital computers, regardless of whether it is possible to construct a perfect continuous perspex machine. However, such a machine would, by its very nature, pass beyond the bounds of language and, presumably, our ability to test it.

I suggest that language is not the hallmark of intelligence, but a restricted case of reasoning and communicating in the continuum.

In the chapter *Visual Consciousness*, I set out the argument that visual consciousness is a partial bi-directional relationship between perspexes. As it stands, this is a functional definition of consciousness, but a robot's body can provide the physical contents of consciousness, as indicated here, and discussed in greater detail in later chapters.

I argue against the notion that there can be any non-physical contents to consciousness, by developing a materialist thesis. I also argue that the supposed non-spatiality of mind is no hindrance to spatial things possessing mind because we can always ignore the supposedly unwanted, spatial properties.

In the chapter *Free Will*, I show how perspexes can be actions and selections. I show how sensorimotor sentences can be conscious as described in the chapter *Visual Consciousness*, and show that this is sufficient to define will as a conscious choice of actions. In defining will in this way, I ignore the question of whether actions are "deliberate." This would make a being responsible for them, and introduce a concept of moral will. Thus the sort of will that is described by perspexes is a very primitive sort of will that might later be developed to deal with the kinds of will we want for ourselves. I then show how a robot can convert its sensations of the world into programs, so that it has a source of programs independent of any external programmer. This also ensures that a robot has individuality, making it difficult for an external agent to coerce a robot, unless the agent knows the robot's individual nature very well. This would seem to be the sort of free will we have. However, there is a severe risk that this kind of free will will make a robot insane, unless its intelligence is sufficient to constrain the effects of executing arbitrary programs absorbed from the world.

In the chapter *Intelligence*, I show that when a Turing or perspex program searches for symmetry it can perform all Turing or, as the case may be, perspex computable actions, so intelligence can be seen as the search for symmetry. More specifically, I show that symmetry describes all repeatable actions which, I suppose, is just about the only kind of action that an animal performs. Repeatable actions make the world predictable, so symmetry describes much of what intelligence does for an animal. Symmetry is also involved in data compression, so it can make programs efficient. Symmetry explains the kinds of data compression that neurophysiologists propose goes on in animal brains. I argue that thinking about the world in any way involves a symmetry between the perspexes that describe a brain and its functioning, and the perspexes that describe the geometrical structure of the world and its functioning. I claim that intelligence is perspex symmetry.

In the chapter *Feeling*, I argue that computers already have functional feelings and already have physical contents to feelings, such as the timeness of time, and the redness of red. I claim that feelings can be intrinsically ineffable, by being Turing-incomputable, and contingently ineffable by being incapable of being shared with some, specific, kind of being. I deny that feelings of any kind are intrinsically incapable of being shared. My identical twin, or myself on a later occasion, might experience what I have experienced, and is now recorded and replayed on a sufficiently competent neurophysiological recording device. I claim that perspexes can have feelings – because they can be the physical body and physical mind of a robot, or of a human, or of any physical thing.

In the chapter *Time*, I review how Turing time is encoded in a perspex machine via the jump part of a perspex instruction. I also describe the more sophisticated model of perspex time. Perspex time has a continuous component of time, as well as a discrete component of Turing time. I describe some of the functional feelings of time that arise from the possession of particular time co-ordinates and describe some of the physical contents of the feeling of time. Some of these relate to changes that occur in the world in general, and some relate to physical properties of a compressed data format that looses detail. Such records are less energetic than the original sensory records. I also propose an experiment to test one way of linking perspex time to physical time. If this hypothesis is correct then it will be possible to construct all manner of devices that exploit time travel. If it is false, no great harm is done. An android should be able to survive perfectly well with a naïve model of time, such as Turing time. After all, this is the model of time that we all use in our daily lives.

In the chapter *Spirituality*, I show how God might be physically real, and possessed of the properties that Christianity claims. I show how spirituality is a manifest property of perspex robots that arises from the symmetry of applying purposes to causes and events in nature. This is entirely rational, has survival value, and is a consequence of being possessed of sufficient intelligence. I show too, that robots are likely to follow our example. To the extent that we engage in cruelty, crime, and war we set a bad example. I argue that if we take on the scientific challenge of creating robots like us – androids – then we should accept responsibility for the traits we include in them. In short, we are responsible for their spirituality and for setting a good example.

Hypothesis

The scientific hypothesis set out in this book is that a very simple thing, the perspex machine, has more computational power than the Turing machine, and can compute any physical phenomenon in the universe. In particular it can describe the geometrical structure and motion of anything. Whilst the Perspex machine almost certainly does not work in the same way as biological neurons, there is a simple interpretation of perspexes, as artificial neurons, which leads to a physical explanation of how a perspex brain can have psychological states of mind. It is hypothesised that a partial bi-directional relationship between perspex neurons manifests visual consciousness; and that the physical universe provides the physical content, or subjective feeling, of which a perspex machine is conscious. It is hypothesised that the translation of perspex visual perceptions into perspex programs provides free will, and a unique, individual identity for a robot. It is hypothesised that the imposition of symmetry in a perspex brain provides intelligence for a robot. It is hypothesised that the perspex machine provides a model of time in which actions can take place simultaneously and reversibly, or else sequentially and non-reversibly. It is hypothesised that physical time works in this way, and that this might be demonstrated by a particular experiment. It is hypothesised that intelligence leads to spirituality and to a conception of God.

Many scientists choose to work on a single hypothesis, but I prefer to carry out research in a more adventurous, multidisciplinary, way. The perspex is interchangeable between objects, shapes, actions, neurons, programs, feelings, intelligence, perception, consciousness, and many other things. We can implement them all in an android and study their interaction. Constraints on any of these things migrate into constraints on all others. This makes a perspex machine very complex, but it also means that every aspect of its being is well matched to the physical circumstances of its existence. Scientific experiments on perspex robots provide the profound engineering challenge of matching the structure and control of a robot's body to the structure and capacities of its brain. The perspex might provide an answer to the grand challenge of making robots that are like us.

Perspex machines can also be studied mathematically. Constraints migrate between perspexes because perspexes maintain homomorphisms, particular kinds of structure preserving transformations, between themselves. The physical and mental aspects of perspex machines are open to a unifying mathematical analysis. We could engage in a mathematical study of the homomorphous unification of machine experience.

All of the above hypotheses exist in themselves. They are described in more detail in the preceding chapters, but there is no need to consider the arguments that lead up to the hypotheses as being part of the hypotheses. The hypotheses are testable in themselves.

Of course, this book does contain philosophical arguments. These are intended to give the reader sufficient confidence to invest the effort needed to carry out scientific tests of the hypotheses, or to allow others to do so.

The central, philosophical argument is a materialistic one: that everything in the universe, including mental phenomena, and God, can be described in physical terems. There is nothing that lies, in principle, beyond the reach of science. This does not diminish the spiritual domain, on the contrary, it elevates the scientific domain to a spiritual one. If we succeed in creating androids that have minds like ours, we will have faced many spiritual challenges along the way, and will face many more in a world enriched by our labours.

Manifestation

The argument set out in this book might seem overblown. How, you might ask, can a scientist imagine that such a simple thing as a perspex can describe such a complicated thing as the universe, and be such a complicated thing as an android's mind? Philosophers might criticise this view by bringing more detail to bear, especially on questions of mind; but they must do more than this if they are to overturn the hypothesis. They must show that, in principle, the perspex cannot do what it is hypothesised to do. They might succeed in this, but it is unlikely that they will invest much effort in this question, unless and until scientific demonstrations begin to appear. The extent to which the hypothesis is true, if at all, will be settled by testing the scientific hypothesis set out in the previous section, perhaps with philosophical criticism to validate the scientific evidence and to circumscribe the bounds of its applicability.

Here I engage in a simpler task, to explain what the scientific paradigm of manifestation is so that it is open to criticism and available for adoption by other scientists. The nearest widely used paradigm is that of genetic algorithms. In that paradigm we set up some complicated system, such as a robot, and provide a very simple description of it in terms of a genetic code with a very few genes. Then we arrange some means of shuffling the genes to produce a variety of individual robots. We then assess how fit each robot is according to some algorithm and arrange that the fitter robots have a greater chance of passing on their characteristics to the next generation. This gives rise to a population with many very fit individuals, and a few less fit ones who, together, preserve the diversity of the gene pool, making the whole population more resistant to extinction. The paradigm of genetic algorithms is very much like biological evolution. Manifestation is like this, but it starts in a very much simpler way.

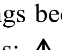
In the paradigm of manifestation we set up a model of a universe. I chose the hypothetical universe of visual robots with a mental life similar to our own and living with us in our universe. We then find the simplest thing which can be everything in the model, and do everything that occurs in the model. I created the perspex to be the simplest thing that can be: a robot's body and mind; a means of linguistic communication with us; an explanation of us in terms of our bodies and the neurophysiology of our brains; and an explanation of the geometrical structure and motion of the universe. The perspex dissolves the mind-body problem in the hypothetical universe by being both things. In the hypothetical universe, things bounce around according to Newtonian physics. The force that drives a robot's mind is twofold. Firstly, a robot responds, entirely causally, to its perceptions of the universe by following its programs. Secondly, it develops and applies motivations, entirely causally, by following programs which seek to impose symmetry in the android's brain and, thereby, in its interpretation of the universe. This preference for symmetry will drive androids to build symmetrical things in the universe such as rectilinear houses, streets laid out in square grids, and space rockets that are rotationally symmetrical. Whatever hypothetical universe we are working in, we must explain its operation in the simplest way, as well as its structure. Conventional mathematics, which explains only structure, is not enough; we must marry this to computer science, or physics, to provide a motive force in the hypothetical universe.

Turing's hypothesis is an example of manifestation. He started by defining the properties of a symbol. This was all he needed to do, but, for the benefit of lesser minds, he went on to define a machine that uses the symbol. The Turing machine is then defined in so much detail that it can be implemented in a physical mechanism, a digital computer, that does not have any mind, other than that created by the Turing machine. Thus, to the extent that digital computers have minds, their minds are manifestations of the Turing symbol. Having *defined* a mathematical model of a computer, Turing then went one stage further. He *hypothesised* that all computers in the physical universe are like the mathematical model. This hypothesis is open to question and, if it is ever shown to be false, we will gain more powerful computers as a result. Thus, Turing set out a purely mathematical definition and applied it, as a scientific hypothesis, to the physical universe. Mathematicians are not so pure as they would sometimes have us believe.

The scientist hypothesises that the manifest model universe explains our physical universe, and accepts the challenge of finding out to what extent the hypothesis is true, and whether it can be modified to make it a more accurate explanation of the physical universe. In this the scientist is open to easy criticism that the model is not so sophisticated as the real universe, but scientists handle this criticism by saying that they investigate models of the real world as a way of making practical progress. Philosophers extend the courtesy of criticising ideas in their own terms so no personal conflict need arise from the adopting this simplification, and some technological progress might be made.

But there is a bold assumption at the heart of the paradigm of manifestation. It is assumed that the simplest thing is sufficient to explain all the phenomena in the hypothetical universe. For example, you might agree that a partial bi-directional relationship between perspexes is necessary to explain visual consciousness, or that symmetry is part of intelligence, but few people, other than myself, would say that this is sufficient; that it provides a complete explanation of visual consciousness and intelligence. I make this claim in the knowledge that the manifest universe is closed. Everything that is, and everything that can be done, in the hypothetical universe manifests from the simplest thing. To say that there is some other property of consciousness or intelligence that is not explained by the simplest thing would be to inject a supernatural element into the hypothetical universe. That is why I extended the discussion of the perspex beyond the four introductory chapters, by including psychological things, and a discussion of spirituality and God. I wanted to indicate that nothing is left out of the hypothetical universe.

It takes a certain chutzpah to wield the paradigm of manifestation – one must explain everything. The slightest omission or error risks bringing the whole hypothesis down. This, according to the principle that scientific hypotheses should be falsifiable, makes manifest hypotheses extremely scientific, because they are extremely falsifiable. But if one has chutzpah, one has it in spades. I hypothesise that the hypothetical perspex universe explains the real one we live in.

I am not, however, entirely reckless. I explained the nature of error, in terms of the walnut cake theorem, so that I know it to be a manifest property of the hypothetical perspex universe. Errors can exist, without damaging the fabric of this universe. In fact, errors are almost inevitable in this hypothetical perspex universe. This confirms our experience of the universe we live in. If I have made some foolish error in my philosophical analysis, this is only to be expected. Greater minds than mine might correct these errors, and to them I leave this task. I will now work on the scientific hypothesis, and put aside philosophical things because I have no further use for them. What more could I possibly say, than this:  ?

APPENDIX 1

Glossary

action

$$\vec{x}\vec{y} \rightarrow \vec{z}$$

$$\text{jump}(\vec{z}_{11}, t)$$

Introduction

This glossary does two things. Firstly, it glosses uncommon words in terms of more common ones, in an attempt to explain the meaning of the uncommon, glossed word to someone with a general grasp of English. The glossed words are also listed in the index, so the reader can find a fuller explanation of them in the body of the book. The reader may also look up the common words in an English dictionary and thereby work out what the uncommon, glossed words mean. However, all such definitions of words are in terms of other words, setting up a vicious circle of words.

One way for people to break out of this vicious circle of words is to refer to bodily sensations. For example, the word, “pain,” is defined in words, but a suitably motivated teacher can usually demonstrate what *pain* is to a student who professes doubt.

The second thing the glossary does is to try to break out of the vicious circle by explaining some basic concepts in terms of the perspex instruction. The perspex instruction describes various physical things, such as light shining through a pin-hole. If we build robots using light shining through a pin-hole, or something equivalent, as their computing element, then they will have all of the computing ability necessary to understand the definition of words in terms of perspexes. We will, however, also want to give robots perceptual abilities similar to ours so that we can agree the meaning of words.

Thus the definitions of words in terms of the perspex shows us how to escape the vicious circle of words. If *we* want to escape, we must understand the physics of light shining through a pin-hole, using our experience of the world and our own mental computing. If we want *robots* to escape, we must build them with sufficient computing power and perception to understand the physics of light shining through a pin-hole, or sufficient awareness of their own perspex computations.

Definitions

action $\vec{\vec{x}}\vec{\vec{y}} \rightarrow \vec{\vec{z}}$; $\text{jump}(\vec{\vec{z}}_{11}, t)$.

afferent The afferent *vectors* are x and y . The afferent *perspexes* are $\vec{\vec{x}}$ and $\vec{\vec{y}}$. They are called afferent by analogy with an afferent nerve that brings information in from the body to the brain. Compare with *efferent* and *transferent*.

affine motion, a *general linear* motion.

Cartesian co-ordinates See *co-ordinates*.

Cartesian space is a mathematical space described by all *Cartesian co-ordinates.* It is also known as "Euclidean space."

computable numbers Numbers which are computable by a Turing machine. All computable numbers are enscribable. See also *enscribable numbers, incomputable numbers*, *semi-computable numbers,* and *number.*

co-ordinates are ordered collections of *numbers* that represent positions in space. **Cartesian co-ordinates** have one co-ordinate for each dimension, or axis, of space. So (x, y) are Cartesian co-ordinates of a 2D space and (x, y, z) are Cartesian co-ordinates of a 3D space. After a while we runs out of letters, so numerical subscripts are used to denote the axes, a , of space. Here (a_1, a_2, a_3, a_4) are Cartesian co-ordinates of a 4D space. **Homogeneous co-ordinates** describe positions in homogeneous space, within which specific ratios of homogeneous co-ordinates describe Cartesian space. The ratios use an extra number, so a Cartesian space of dimension D is described by a $D + 1$ homogeneous space. The extra dimension is needed to provide the extra number. Homogeneous spaces support a simpler description of perspective than Cartesian spaces, which is why they are used in computer graphics and robot vision.

counter A fragment of program, or a piece of hardware, that counts by adding one to a variable. A counter can be stopped, or re-set to zero, but otherwise it just counts. Counters are used to make clocks or to count the number of events that occur. For example, a counter might count the number of key strokes a user types at a computer key-board. Counters also play a theoretical role in the theory of computability.

Delaunay triangulation A particular division of a shape into compact triangles or, more generally, into *simplexes*.

Dendrites Data and control paths from the cell body of a perspex neuron to the points x, y, z, and four positions of t.

efferent The efferent *vector* is z. The efferent *perspex* is $\vec{z}$. They are called efferent by analogy with an efferent nerve that takes information outward from the brain to the body. Compare with *afferent* and *transferent*.

enscribable numbers Numbers which can be written as a decimal expansion on the data tape of a Turing machine. Enscribable numbers can be computable, incomputable, or semi-computable. See also *computable numbers, incomputable numbers, semi-computable numbers,* and *number.*

field In physics a *field* is a spatial distribution of some kind of thing. A *scalar field* is a distribution of some numerical quantity, such as temperature or pressure. A *vector field* is a distribution of some vector quantity, such as a gravitational field. There are other kinds of field. Perspex program space is a *perspex field*. To the extent that an integer is non-spatial in itself, an *integer field* is a spatial distribution of non-spatial things – the integers. Mathematics also defines *fields,* but they are particular kinds of algebraic structures.

general linear motion, *linear* motions and *translation*.

homogeneous co-ordinates See *co-ordinates*.

homogeneous space A mathematical space described by all *homogeneous co-ordinates.* Compare with *perspective space* and *program space*.

homomorphism A specific kind of mathematical transformation which preserves the structure of the thing being transformed, but which cannot, in general, be undone. Homomorphisms which can be undone are called isomorphisms.

incomputable numbers Numbers which a Turing machine cannot compute. For example, all irrational numbers with infinitely many random digits are incomputable. However, the theoretical perspex machine can, by definition, carry out computations with any given number, even if it is incomputable by a Turing machine. For example, all numbers on a segment of the number line, including Turing-incomputable numbers, can be given to a pin-hole, perspex machine and can be operated on in a single operation. A pin-hole, perspex machine can also carry out incomputable operations of “arithmetic” and “logic.” See *computable number, enscribable number, semi-computable numbers,* and *number.*

integer numbers 0, ± 1, ± 2, ± 3, $\ldots$

irrational numbers Numbers which are not rational or transrational. For example: $\sqrt{2}$, π, and e.

linear motion, a motion containing any or all of the motions: shear, rotation, scale (magnitude), and reflection (handedness).

manifold A surface with specific geometrical and topological properties that makes it well behaved. Manifolds can be used for many things. In computer vision they define functions that show how close any seen shape is to the prototypical shape defined by the manifold. In this book manifolds are used to define the nearest word to a continuously varying meaning.

motion (perspex) $\vec{x}\vec{y} \to \vec{z}$. See *transformation*.

normalisation Putting into a standard from.

number The individual elements x_i or y_i or z_i or t_i. Compare with *vector*. See also *computable numbers, enscribable numbers, incomputable numbers, integer numbers, irrational numbers, rational number*s, *real numbers, semi-computable numbers,* and *transrational numbers*.

number line A geometrical line that contains all of the real numbers.

origin The point where co-ordinate frames start, that is, the point where they, "originate" from. The origin of 2D and 3D Cartesian co-ordinate frames is, respectively, $(0, 0)$ and $(0, 0, 0)$. The point $(0, 0, 0)$ is also the origin of a 3D homogeneous co-ordinate frame that describes 2D Cartesian points. The 2D Cartesian origin $(0, 0)$ is $(0, 0, 1)$ in 3D homogeneous co-ordinates.

perspective A geometrical distortion that describes the shape of an object as it is seen in a pin-hole or thin-lens camera.

perspective space A mathematical space described by all *homogeneous co-ordinates,* except the homogeneous *origin.* Compare with *homogeneous space* and *program space*.

perspex A portmanteau of the words "perspective" and "simplex", which originally referred to a simplex of dimension D embedded in a homogeneous space of dimension $D + 1$. Now specialised to a $3D$ simplex embedded in a $4D$ program space.

pin-hole The simplest way to produce a perspective image is to view the world through a pin-hole.

program space A mathematical space described by all *homogeneous co-ordinates.* Every point in this space contains a *perspex*. Compare with *homogeneous space* and *perspective space*.

projection See *transformation*.

quaternion A specific kind of mathematical object containing four numbers. Quaternions can describe positions in space and *general linear motions*. They are most commonly used to describe rotation.

rational numbers $\frac{0}{1}, \frac{\pm 1}{1}, \frac{\pm 1}{2}, \frac{\pm 1}{3}, \frac{\pm 2}{3}, \frac{\pm 1}{4}, \frac{\pm 3}{4}, \ldots$

real numbers The rational and irrational numbers.

selection $\text{jump}(\overset{\geq}{z}_{11}, t)$.

semi-computable numbers Numbers for which only an upper or else a lower bound is computable by a Turing machine. In some cases the series of computed bounds is so well behaved that a Turing machine can compute the semi-computable number. If both an upper and a lower bound is computable then the number is computable. If no bounds are computable, or the sequence of computable bounds is badly behaved, then the number is incomputable. See *computable numbers, enscribable numbers, incomputable numbers,* and *number.*

simplex The simplest, straight-edged figure that contains a volume of the space it is embedded in. In 3D a tetrahedron, in 2D a triangle, in 1D a line, in 0D a point. Higher dimensional simplexes exist, but geometrical figures, such as simplexes, do not exist in spaces of dimension less than zero.

spacetime an arrangement of several spatial dimensions and one time dimension.

synapse The location where an *afferent*, *efferent*, or *transferent dendrite* meets the body of a perspex neuron.

transferent Relating to any, or all four, control paths used by $\text{jump}(\overset{\geq}{z}_{11}, t)$ that transfer control from the body of one perspex neuron to another. Compare with *afferent* and *efferent*.

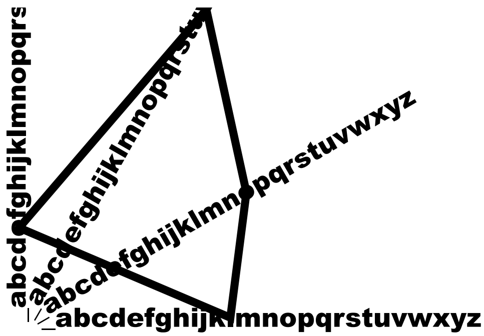
translation A geometrical change of position, that is, a change of *origin*. A change between equivalent expressions in a language or languages. A change between equivalent elements in the continuum. These may be computable or incomputable numbers. In the latter case the translation is super-linguistic.

transformation A change of some kind. A geometrical *motion*. Non-singular transformations preserve dimensionality. Singular, transformations are called *projections*, these reduce dimensionality. The *transformation* from one 4D perspective space to another 4D perspective space is non-singular, but the *projection* from a 4D perspective space onto a 2D picture is singular.

transrational numbers infinity, $\infty = 1/0$, and nullity, $\Phi = 0/0$.

vector The ordered collection of numbers x_i or y_i or z_i or t_i.

APPENDIX 2 *Visual Phrases*



Introduction

A number of English phrases refer to the sensory modality of vision. Here we give an alphabetical list of some of these visual phrases. For human speakers of English these phrases have actual and/or metaphorical meanings, but for a perspex machine many of the metaphorical meanings might be actual. Thus the list gives possible points of contact between humans and perspex robots where each kind of being might see how language refers to the mental structures of the other. The perspex structure of a phrase can be regarded as containing the "deep structure" alluded to in a Chomskian grammar of human language. The perspex structure is also capable of recording and carrying out the transformations of any transformational grammar.

The list is drawn up under headings such as "Blind," "Eye," "Look," and "See." Phrases are placed according to the first head word in the phrase. Thus, "to see with one's own eyes" is placed under "see" and not "eye." An alternative which may be present or not is shown in round brackets. Thus "(34) poke in the eye (with a sharp stick)" stands for the two phrases "(34.1) poke in the eye" and "(34.2) poke in the eye with a sharp stick." If exactly one of the alternatives must be used, the alternatives are put in square brackets and are separated by vertical bars. Thus "(9) blinding [light | pain]" stands for the two phrases "(9.1) blinding light" and "(9.2) blinding pain." Where necessary the alternatives within a numbered phrase can be further sub-numbered. Thus, "(87) see through [a glass darkly | rose tinted [glasses | spectacles] | the bottom of a glass | the [vail | veil] of tears]" yields two phrases, amongst others, "(87.2.1) see through rose tinted glasses" and "(87.2.2) see through rose tinted spectacles."

The lists were composed from the author's own memory and from the compendia listed under *Further Reading*. The phrases were checked by doing a word search with a world-wide-web search engine. There is no attempt to draw up an exhaustive list of visual phrases.

Alphabetical List

Behold

1. behold

Bleary

2. bleary (eyed)
3. Blind
4. [bake | body | colour | gravel | fight | night | play | sand | snow | stark | word] blind
5. blind
6. blind [as a bat | to the consequences]
7. blind [hope | faith | justice | side | sight | spot | watchmaker]
8. blinding
9. blinding [light | pain]
10. blindingly
11. blindingly [obvious | stupid]
12. blindfolded

Blinkered

13. blinkered

Eye

14. an eye for an eye
15. an eye for the [boys | girls | main chance]

16. [black | blue | brown | bull’s | eagle | evil | gimlet | green | grey | hawk | hazel | jaundiced | lynx | mind’s | naked | pink | red | sharp | swivel | unaided | x-ray] [eye | eyes]
17. [blue | cross | eagle | green | hawk | lynx | one | sharp | skew | wall] eyed
18. clap eyes on
19. eyeball
20. eye-catching
21. eye of [a needle | faith]
22. eye [strain | test | witness]
23. eyeful
24. eyeless
25. eyes right
26. four eyes
27. get one’s eye in
28. hit between the eyes
29. in the [blink of an eye | eye of a storm | wink of an eye]
30. keep half an eye on
31. meet the eye
32. [glint | one] in the eye
33. out of the corner of the eye
34. poke in the eye (with a sharp stick)
35. rivet one’s eyes
36. use your eyes
37. within eyesight

Gape

38. gape

Gawk

39. gawk

Gawp

40. gawp

Glance

41. exchange glances
42. glance
43. [furtive | inviting | sideways | quick] glance

Glare

44. glare

Goggle

45. goggle

Hoodwink

46. hoodwink

Illusion

47. illusion

Invigilate

48. invigilate

Leer

49. leer

Look

50. dirty look
51. look
52. look [askance | here | see]
53. look and learn
54. look at it my way
55. look at it from my [point of view | stand point]
56. look out (for oneself)

57. look [over | through]

58. look to the [future | past]

59. look to your [family | honour | left | right | sword]

60. looker

61. on looker

62. outlook

Make Out

63. make out

Observe

64. observe

Ogle

65. ogle

Panorama

66. panorama

Peek

67. peek(aboo)

Peep

68. peep

Perspective

69. perspective

70. [from my | keep a | oblique] perspective

Perspicacious

71. perspicacious

Pore Over

72. pore over

Prospect

73. prospect
74. [beautiful | exciting | fine | good | high | low] prospect

Regard

75. regard
76. [high | low | no] regard

Review

77. review
78. revision

Rubber necking

79. rubber necking

Scan

80. scan

See

81. see (here)
82. see beyond the end of your nose
83. see into the seeds of time
84. see it [my way | through (to the end) | to the end]
85. see my way clear
86. see [red | Rome and die | the world]
87. see through [a glass darkly | rose tinted [glasses | spectacles] | the bottom of a glass | the [vail | veil] of tears]
88. see to the heart of the matter
89. see with one’s own eyes

90. sight-see
91. unseeing

Scene

92. scene
93. make a scene

Scowl

94. scowl

Show

95. Show
96. Show through

Sight

97. A sight for sore eyes.
98. [catch | dim | failing | first | long | near | partial | second | short | weak] sight
99. in plain sight
100. sight
101. sightless
102. oversight

Stand Point

In the pre-computer age, perspective drawings were constructed using a *stand point,* that is, the point on the ground plane where the viewer stands, and a *view point* at the height of the viewer's eyes directly above the stand point. For convenience a height of five feet was often used. I do not know whether the technical or common use of these words was historically prior, but this would seem to be a straight forward question of etymology.

103. from were I stand
104. my point of view
105. points of view
106. share my (point of) view

Stare

107. stare

108. stare [at | into one's eyes | longingly | lovingly | past]

Survey

109. survey

Squint

110. squint

View

111. view

112. [beautiful | bird's eye | dramatic | ghastly | horrible | jaundiced | skewed | upsetting | worm's eye] view

Vigil

113. vigil

Visible

114. (in)visible

Vision

115. [binocular | double | monocular | stereoscopic] vision

116. vision

117. visionary

118. visionless

Vista

119. vista

Visualise

120. visualise

Watch

121. keep watch

122. watch

123. watch [out | your backs]

124. watch and [hope | pray]

125. watch [my lips | the clock | the pot boil | with mother]

Witness

126. witness

127. bear false witness

Further Reading

1. *Collins English Dictionary* Harper Collins (1991).
2. *Roget's Thesaurus of English Words and Phrases*, Penguin Books, 1966. First published 1852.

APPENDIX 3

Annotated Bibliography

$$\text{jump}(\vec{z}_{11}, t)$$

3. Alexander, S. *Space, Time, Deity* Macmillan, (1927). First printed 1920.

This metaphysical treatise is aimed at the general reader and the professional philosopher. The central argument is that everything in the universe – including all mental phenomena, such as feeling, free will, consciousness, and God – is explicable in terms of spacetime. The argument is very long, being set out in two volumes, and relates each of a series of physical, mental, and spiritual, phenomena to spacetime. Each relationship is set out in its own terms, without any specific, unifying principle.

The argument for God is unusual. Alexander argues that it is our experience of the universe that things develop into more complex forms, that the most complex form we currently know of is mind, but that we may reasonably suppose that something more complex than mind will develop in the universe. Alexander calls this next more complex thing *deity* and says that a being in possession of deity is *a* god or, in the case of one, *the* God. Alexander argues that this state of the universe has not yet come about, but that spiritual practices are motivated by our appreciation of whatever it is that is developing into deity.

There is no bibliography, but there are footnotes giving references to a small number of philosophical and scientific works. Many of the scientific works are now of only historical interest, but in the second, 1927, impression he refers to a number of published critiques of the first, 1920, impression

4. **Altmann, S.L.** *Rotations, Quaternions, and Double Groups* Clarendon Press, Oxford, (1986).

This mathematical treatise is aimed at the advanced student of physics and chemistry who is familiar with group theory and calculus. The book describes 2D, 3D, and multidimensional rotations, mainly, in terms of quaternions. There are worked examples, many of them demanding. The bibliography refers mainly to research papers in the physical sciences and mathematics.

5. **Anderson, J.A.D.W.** "Visual Conviction" Proceedings of the fifth Alvey Vision Conference, University of Reading, England, pp. 301-303, (1989).

This philosophical paper is aimed at the general reader with an interest in computer vision. It argues that the essence of visual knowledge that distinguishes it from other kinds of knowledge is that it is knowledge which is in a bi-directional mapping with an image. There is also a discussion of the kinds of knowledge about itself that a fallible mind can have. It is concluded that a fallible mind can have, at best, consistently justified beliefs about itself. Such beliefs are called, "convictions" and are visual when they are in a bidirectional mapping with an image.

6. **Anderson, J.A.D.W.** *Canonical Description of the Perspective Transformations* Ph.D. Thesis, Department of Computer Science, the University of Reading, England, RG6 6AY, (1992).

This Ph.D. thesis gives a mathematical derivation and program code for generating all affine transformations, of any dimension, from standard parameters, and for combining these transformations with a single focal-ratio, perspective transformation. The inverse of this procedure is also given – it returns the standard parameters of a transformation.

The thesis analyses and clears up a confusion about the terms *skew* and *shear* as these terms were used in the contemporary computer graphics literature. This literature now correctly uses the term *shear* to refer to either a strictly upper, or else strictly lower, triangular, matrix transformation. The term *skew* properly refers to lines in 3D or higher dimensions that do not intersect. However it continues to be used to mean a 2D rotation, or a 2D orthogonal projection of a 3D rotation, in phrases such as, "de-skewing text prior to optical character recognition." It would be more accurate and clearer to say, "de-rotating text prior to optical character recognition." In optical character recognition lines are de-rotated and hand written strokes within a line are de-sheared. These are different operations, but if you want to refer to both, why not say, "text is geometrically normalised prior to optical character recognition." This would

also encompass transforming text into a standard size which is, usually, what goes on in the optical recognition of hand written text.

The bibliography refers to scientific papers and text books in computer graphics

7. **Anderson, J.A.D.W.** "Representing Visual Knowledge" in *Philosophical Transactions of the Royal Society* vol. 352, pp. 1129-1139, (1997).

This scientific paper is aimed at the AI researcher. It introduces the point at nullity, the first version of the perspex, and has a discussion of time that is now completely superseded. The results of one computer vision program using this early version of the perspex are shown.

8. **Anderson, J.A.D.W.** "Perspex Machine" in *Vision Geometry X1*, Longin Jan Lateki, David M. Mount, Angela Y. Wu, Editors, *Proceedings of SPIE* Vol. 4794, 10-21, (2002).

This paper is aimed at the mathematician familiar with projective geometry and the theory of computability. It shows how to use perspective transformations to implement a Turing machine. It is hypothesised that this could lead to the development of very fast, optical computers. Enough detail is given to implement a model of a perspex machine using standard computing techniques.

It is shown that the perspex machine is irreversible in time, leading to a temporally anisotropic spacetime. The extreme hypothesis is made that time in the physical universe operates in this way and an experiment is proposed to test this hypothesis using the Casimir apparatus.

9. **Anderson, J.A.D.W.** "Exact Numerical Computation of the General Linear Transformations" in *Vision Geometry X1*, Longin Jan Lateki, David M. Mount, Angela Y. Wu, Editors, *Proceedings of SPIE* Vol. 4794, 22-28, (2002).

This paper is aimed at the student who is familiar with trigonometry. It parameterises all of the rational rotations in terms of a transrational number which is equal to the half tangent of the angle of rotation. In addition to the rational numbers there are two strictly transrational numbers – infinity, $\infty = 1/0$, and nullity, $\Phi = 0/0$. Infinity denotes a positive rotation by a right angle and nullity denotes all of the trigonometric ratios of a degenerate triangle with sides of zero length.

The practical consequence of this is that rotation, and all affine and perspective transformations, can be computed without the risk of arithmetic underflow or overflow. Hence very robust computer vision programs can be implemented.

There is, however, an omission in the paper. The sign convention in equation 18 is not explicit. The convention is in two parts. Firstly, the integer square root is signed. That is, the positive or negative root, $\sqrt{x}$, is chosen so that $\operatorname{sgn}(\sqrt{x}) = \operatorname{sgn}(x)$. Secondly, the radius, r, is non-negative. Consequently the sign of the denominators p and q of r/p and r/q is chosen so that $\operatorname{sgn}(p) = \operatorname{sgn}(r/p)$ and $\operatorname{sgn}(q) = \operatorname{sgn}(r/q)$.

10. **Anderson, J.A.D.W.** "Robot Free Will" in *ECAI 2002 Proceedings of the 15th European Conference on Artificial Intelligence,* Lyon, France, ed. F. van Harmelan, pp. 559-563, (2002).

This paper provides a simple summary of the perspex neuron and shows how the perspex may be used to give robots a will that is free of coercion by others, including being free of their original, human, programmer.

11. Audi, R., editor, *The Cambridge Dictionary of Philosophy*, Cambridge University Press, (1995).

A good, detailed dictionary of philosophy. The many articles provide an overview of philosophical works sufficient to direct the general reader to criticise the primary sources. However there is no bibliography of these sources.

12. **Ayres, F.** *Schaum's Outline Series: Theory and Problems of Matrices* McGraw Hill, (1974).

This text book aims to teach matrix algebra to students with a modest mathematical background. It has very many worked examples at various levels of difficulty. There is no bibliography.

13. **Ballard, D.H. & Brown, C.M.** *Computer Vision* Prentice-Hall, (1982).

This text book aims to review the main techniques used in computer vision at its publication date and to present these to programmers who are familiar with statistics, vector algebra, and calculus. Each chapter has a bibliography of scientific papers and text books in computer graphics.

14. **Banchoff, T.F.** *Beyond the Third Dimension* W.H Freeman, (1990).

This heavily illustrated book is aimed at the general reader who knows how to count and wants to understand spatial dimensions and especially spaces with more than three dimensions. The book introduces mathematical ideas in a non-technical way and illustrates these with many colour diagrams and photographs

of the natural world. A short bibliography cites technical magazine articles, books, and sources of illustrations, all of interest to the general reader.

15. **Boden, M.** *Purposive Explanation in Psychology* Harvester Press, (1978). First published 1972.

This monograph, with an extensive bibliography, compares a number of historical theories about the nature of purpose in psychology and philosophy. Boden's conclusion is that purposes are reducible to mechanistic processes, but that it is useful to talk of purposes for three reasons. Firstly, purposes are accessible to a mind and so can be thought about and communicated to others, whereas some of the myriad bodily and physical circumstances that make up a purpose are inaccessible. Secondly, the myriad detail is too voluminous and complex to be thought about, whereas purposes provide a short hand which can be thought about and communicated. Thirdly, purposes can apply to very many different complexes of bodily and physical circumstances, so they allow a reasoning by analogy from one occasion to another or from one mind to another. Boden concludes that purposes are reducible to mechanistic processes, but that the high level of purpose is useful in itself.

16. **Boden, M.** *The Creative Mind* Cardinal, (1992). First published 1990.

This highly readable book sets out the view that a computer can represent problems in an abstract "space" and can be creative by combining elements of solutions in new ways. Boden gives some guidance on what sorts of combinations typically lead to useful creativity. She draws a distinction between being personally creative, thinking of something new for oneself, and being historically creative, thinking of something that no one else has yet thought of. Being historically creative is simply a contingent fact of history. The same mechanisms of creativity are involved in both cases. A bibliography contains mainly books and papers in Artificial Intelligence.

17. **Boolos, G.S. & Jeffrey, R.C.** *Computability and Logic* Cambridge University Press, (1996). First Published 1974.

This text book for the Open University is aimed at the Computer Science, Mathematics, or Philosophy student who wants to understand the fundamental limits inherent in Turing computability. The book makes heavy use of formal arithmetic and logic. There are a number of exercises and some partial solutions. The exercises are lengthy, detailed, and occasionally demanding. There is no bibliography.

18. Carroll, L. *Alice's Adventures in Wonderland* and *Through The Looking Glass* Puffin Books, (1987). First published 1865.

These two children's stories have proved enduringly popular. Some of the apparently nonsense sayings and words relate to contemporary mathematics, at both an elementary and advanced level. This adds a dimension of amusement for children and adults with a mathematical education; but the book is pure fun.

The verse about the bat in my chapter *Feeling* appears in Carroll's chapter *A Mad Tea-Party*, of *Alice's Adventures in Wonderland.*

19. Carroll, L. *The Hunting of the Snark* (1876). Reprinted with mathematical annotations in **Gardner, M.** *The Annotated Snark*, Penguin Books (1984). First published 1962.

This nonsense poem has proved as enduringly popular as Alice's Adventures in Wonderland. Martin Gardner provides a mathematical and historical commentary on the allusions in this poem.

The verse quoted in my *Preface* comes from Carroll's *The Beaver's Lesson*. The two verses quoted in my chapter *Time* come from Carroll's *The Vanishing*. I have modified the punctuation of the first of these two verses, so that it can stand, grammatically, without its preceding verse.

20. Chomsky, N. *Syntactic Structures* Mouton, (1957). Reprinted 1985.

Chomsky's original book setting out the principle that all human languages have a common deep structure, embedded in our brains, that is transformed by grammatical re-write rules into the sentences of any language.

21. Churchland, P.M. *The Engine of Reason, the Seat of the Soul: a Philosophical Journey into the Brain* MIT Press, (1996). First printed 1995.

This highly readable book for the general reader introduces the anatomy of the human brain and describes the major, external sensory systems. It puts forward the Parallel Distributed Processing (PDP) model of brain function in which networks of standard, artificial neurons operate on high dimensional vectors. There is a very clear introduction to the philosophy of consciousness. The book comes with a flat-packed stereoscope to view the 3D photographs in the chapter on stereo vision. An eclectic bibliography cites papers and books on the topics of neuroscience, artificial intelligence, and philosophy.

We might ask why we want two separate things – neurons and vectors – when the perspex can be both things, and more.

22. **Clarke, C.J.S.** *"The Nonlocality of Mind"* in **Shear, J.,** editor, *Explaining Consciousness* – 'The Hard Problem', MIT Press, pages 165-175, (1995).

This philosophical essay, which is accessible to the general reader, claims that mind is not located in space, but existed prior to the physical universe. Clarke recommends that physicists accept mind as the most fundamental thing in the universe and develop the properties of quantum physics by using this universal mind as an observer of quantum events.

23. *Collins English Dictionary* Harper Collins, (1991).

24. **Cotogno, P.** "Hypercomputation and the Physical Church-Turing Thesis" *British Journal for the Philosophy of Science,* number 54, pp. 181-223, (2003).

This philosophical paper is aimed at the professional philosopher with an interest in computability. Cotogno reviews several approaches to constructing a super-Turing computer and concludes that there is little prospect that any of them will succeed.

25. **Coxeter, H.S.M.** *Projective Geometry* 2nd edn., Springer-Verlag, (1994).

This text book is aimed at mathematics students. It develops projective geometry via the operations of *join* and *meet*, introduces the principal of *duality*, and presents a number of classical theorems. The *cross-ratio* and *homogeneous coordinates* are presented. A moderate number of worked examples give the student good opportunity to test understanding and develop problem solving skills. A short bibliography cites, mainly, mathematical text books.

26. **Crownover, R.M.** *Introduction to Fractals and Chaos* Jones and Bartlett, (1995).

This text book is aimed at the numerate science student with a background in programming who is familiar with the mathematics of complex variables and calculus.

The book provides a good introduction to self-similar fractals, various measures of fractal dimension, and the relationship between fractals and the chaos that arises in both deterministic and stochastic systems.

There are numerous figures showing fractals, including six, striking colour plates. There are mathematical proofs, program fragments, and suggested exercises. Whilst some of these exercises give strong guidance on their solution, there are no worked solutions. There is a short, but useful bibliography of early work on fractals and chaos.

27. **Cutland, N.J.** *Computability: an introduction to recursive function theory* Cambridge University Press, (1980).

This text book is aimed at the mathematics student who has some facility with conventional programming. It sets out the fundamental limits inherent in Turing computability. The book makes heavy and very effective use of the Unlimited Register Machine (URM) as a more convenient model of computation than the Turing machine and analyses it, almost exclusively, in terms of recursive functions. There are a number of graded exercises, most of which require technical skill in mathematics, and some of which are demanding. No solutions are given explicitly, but they can sometimes be found in the text. A short bibliography cites mathematical papers and text books.

28. **Dennett, D.C.** *Elbow Room* Oxford University Press, (2002). First printed 1984.

This book is aimed at the general reader. It sets out to show that free will is compatible with a deterministic universe. The style of the argument is interesting. First of all Dennett analyses and disposes of many arguments that seek to show that determinism denies a being free will. Of course, Dennett cannot answer all such arguments, so he concludes by giving a recipe for criticising any other arguments against free will, leaving it up to the reader to carry out that criticism. A bibliography cites mainly philosophical text books and papers, though some works in biology and AI are cited.

29. **Dreyfus, H.L.** *What Computers Still Can't Do: a critique of artificial reason* MIT Press, (1997). First published 1972.

This monograph is aimed at the general reader and to practitioners in Artificial Intelligence (AI). It sets out the philosophical thesis that AI must fail for two reasons. Firstly, that AI does not deal with a program embodied in a robot that takes part in human society. Secondly, that digital computers cannot model continuous computations as required, for example, to model graded potentials in the primate retina and analog filtering in neurons. Dreyfus claims that AI practitioners have not given any reason to believe that Turing computation is sufficient to explain human reasoning. He urges the AI community to examine the causes of its failure and to propose deeper models of computation embodied in a robot. A bibliography of books and papers on AI and Philosophy is given as chapter notes.

30. **Farrell, B.A.** "Experience" in **Chappell, V.C.** *The Philosophy of Mind* Prentice-Hall, pages 23-48, (1962). This essay was first published in the journal *Mind* in 1950.

This essay is accessible to the general reader, but is so imbued with rhetorical questions in the form of statements, and an apparently ironic comment on witchcraft, that it is difficult to follow the author's argument. This seems to be that scientists can discover what animal behaviour is, but have no way to describe the experience, or feelings, that animals have. The best a human can achieve is to discover what it is like to be other kinds of human, by taking up their life style. Read in one way, the essay argues that witchcraft might, theoretically, allow a human to be turned into a terrestrial animal, so we could feel what it is like to be a bat, but cannot turn a human into a Martian, so we cannot feel what it is like to be a Martian. The supposed terrestrial nature of witchcraft is not explained. Read in another way, bats are mammals and have biological similarities to us, but Martians, or extraterrestrial beings, we suppose, have a biology so different from us that we have no hope of ever being bodily transformed into them, so we can never experience what it is like to be an extraterrestrial being. Perhaps Farrell's argument would be stronger if he had stopped at the claim that a human, or animal, can only ever experience its own feelings.

The questions of what it would be like to be an opium smoker, bat, or Martian are introduced in pages 34 and 35 of the *Philosophy of Mind.*

31. **Foley, J.D., van Dam, A., Feiner, S.K. & Hughes, J.F.** *Computer Graphics: Principles and Practice* 2nd edition, Addison-Wesley, (1990).

This reference text book provides the most comprehensive introduction and review of computer graphics techniques. It is aimed at strong programmers who are familiar with the mathematical techniques of matrix and vector algebra and calculus. It has a large bibliography of research papers and text books on computer graphics.

32. **Franklin, S.** *Artificial Minds* MIT Press (1995).

This highly readable book reviews a number of AI programs and seeks to show that animals, computers, and simple machines have minds. However, there is no single, unifying principle to this argument. Franklin report's Aaron Sloman's view that there is no sharp distinction to be drawn between possessing free will and not possessing it. Instead there are varying degrees of free will depending on what choices a machine can envisage.

(Having spoken to Aaron on many occasions, I would say that he has a pretty dim view of sharp distinctions between the possession and absence of any mental faculty. By contrast, he has a brilliant view of the diversity of computational mechanisms that might provide such faculties.)

The book has a bibliography, mainly composed of AI papers, with some philosophical and ethological works.

33. **McGinn, C.** *The Problem of Consciousness* Basil Blackwell Ltd., (1991).

This highly readable collection of philosophical essays sets out a pessimistic view of the prospect of understanding consciousness. McGinn argues that whilst consciousness is, by definition, accessible to a mind that possesses it, such a mind cannot have access to the underlying physical processes that give rise to consciousness. Furthermore, that the human mind is, in principle, incapable of understanding consciousness. However, he accepts, in the final essay, that a computer could have consciousness if it were appropriately programmed; but he supposes that it is, probably, beyond human intelligence to write such a program.

34. **McGinn, C.** "Consciousness and Space" in **Shear, J.,** editor, *Explaining Consciousness* – 'The Hard Problem', MIT Press, pages 97-108 (1995).

This philosophical essay claims that brains exist in space, but minds do not. McGinn suggests that understanding consciousness might require such a radical paradigm shift in physics, to reconcile non-spatial minds with spatial brains, that it might be beyond the power of the human mind to understand consciousness.

35. **Nagel, E. & Newman, J.R.** *Gödel's Proof* Routledge, (1993). First published 1958.

This short, highly readable book is aimed at the general reader. It gives an insightful and easy introduction to Gödel's proofs. A very short bibliography cites Gödel's original work, which is extremely difficult, and lists more accessible works on computability.

36. **Nagel, T.** *What is it Like to be a Bat* in **Hofstadter, D.R. & Dennett, D.C.** *The Mind's I* Penguin Books, pp. 391-403, (1986). Book first published 1981. Paper first published in 1974 in the *Philosophical Review.*

Nagel considers the question of what it feels like to be a bat and a Martian. That is, he asks what a bat and a Martian feel. He concludes that there is no way for us to know, because we are not bats or Martians. The most we can hope for, he claims, is to know what the physical correlates of feelings are.

There is no bibliography and only a very few references in footnotes in the text that appears in the book.

37. **Noë, A. & Thompson, E.** *Vision and Mind: Selected Readings in the Philosophy of Perception* MIT Press, (2002).

This is a compendium of psychological and philosophical papers that marked turning points in the philosophical theory of visual perception. There is no attempt to synthesise the different views.

38. **Popper, K.R. & Eccles, J.C.** *The Self and its Brain - an argument for Interactionism* Routledge, (1995). First published 1977.

This long book is aimed at the general reader. It has three parts. In the first part Karl Popper reviews the philosophy of mind and sets out a theory that mental things, ideas, have an abstract reality that allows them to influence the behaviour of humans and hence to change the arrangement of the physical world. He regards the world as being an ecology of physical things and abstract ideas.

In the second part John Eccles describes the neuro anatomy, but very little of the neuro physiology, of the brain. He describes the main sensory pathways and discusses evidence that consciousness lags half a second behind transduction of a physical quantity at the sensory receptors. He reports the suggestion that the brain ante-dates percepts by this time lag. He sets out the view that there is a spiritual mind that interacts with the physical brain.

Chapter E1 contains the claim that there are up to 10 000 neurons in a cortical column. Chapter E3 describes an electrical “readiness potential” that occurs in the motor cortex prior to any “willed” action.

The third part is a series of dialogues between the two authors that deals with many aspects of consciousness in non-human animals, children, and adults. Language and visualisation are discussed, among many other phenomena. The book comes to the standard conclusion that, in humans, language created the abstract world of ideas.

The bibliographies in each of the three sections of the book refer to books and scientific papers in English and German on philosophy, neurophysiology, and psychophysics.

39. **Pour-El, M.B. & Richards, I.J.** *Computability in Analysis and Physics* Springer Verlag, (1989).

This monograph is aimed at the postgraduate student with a strong background in mathematical analysis (the theory underlying calculus) and computability. All of the examples are taken from mathematical physics, so familiarity with this subject makes the examples more helpful. Many theorems are given at varying levels of difficulty, but there are no exercises. A bibliography cites text books and scientific papers on mathematical topics related to the computability of physical functions.

The introductory chapter reviews standard results in computability and analysis, some of which are glossed in my *walnut cake theorem.* The idea of a computable real number as the limit of both monotonically rising and falling computable sequences of rational numbers is glossed as the kernel of truth lying inside one slice of cake. The idea of a real number being the limit of monotonic sequence of computable rational numbers is glossed in the discussion of semi-computable numbers, such that the kernel of truth lies in an unknown number of slices of cake above or else below a cut. In the case that the convergence of this sequence is computable, the real number is computable. The main thrust of the walnut cake theorem is to do with computable numbers, not strictly semi-computable numbers, or incomputable numbers, but as these exist they should be considered.

Theorem 6 in section 5 of chapter 3 discusses the wave propagation equation of physics. It shows that with certain starting conditions and norm there is a unique solution to the equation which is continuous but incomputable. This implies that there are smoothly varying physical phenomena that are incomputable. If such phenomena exist they might be exploited to construct super-Turing machines. But see Weihrauch, K. and Zhong, N[54].

40. **Proudfoot, M.**, editor, *The Philosophy of Body*, Blackwell, (2003).

This short book is a collection of philosophical papers, mainly by professional philosophers, on the role of the body in shaping human perceptions, social and sexual categories. It sets out to reinforce a new movement to discuss the body in philosophy as a counterpoint to an excessive, historical preoccupation with the mind.

The papers assume a familiarity with philosophical literature that can be obtained from introductory texts in philosophy or by consulting a good dictionary of philosophy. See Audi, R. *The Cambridge Dictionary of Philosophy.*

41. **Riesenfeld, R.F.** “Homogeneous Coordinates and Projective Planes in Computer Graphics” *IEEE CG&A* pp. 50-55, (1981).

This short paper, aimed at the numerate scientist, explains the use of homogeneous co-ordinates to describe the affine and perspective transformations. It explains why homogeneous space is unorientable and indicates that clipping can be used to obtain an orientable image from the point of view of a physical camera.

42. **Roget, P.M.** *Roget's Thesaurus of English Words and Phrases*, Penguin Books, (1966). First published 1852.

A standard reference work listing English words and phrases of a similar meaning.

43. **Scruton, R.** *Animal Rights and Wrongs,* Demos, (1996).

This highly readable, short book is aimed at the general reader. It is a political essay informed by philosophical literature and methodology. It proposes that we are entitled to believe that animals do not have minds unless there is evidence to the contrary. Such evidence is part of folk psychology for many kinds of animals. The book further proposes a hierarchy of animal rights and duties depending on the kind of mind an animal possesses. A self-conscious mind, as possessed by the majority of adult humans, engages in a moral web of rights and obligations established by dialogue with each other. Such minds may accept duties toward lower minds such as human infants, mentally disadvantaged adult humans, and other animals. The duties owed vary according to whether a moral mind explicitly accepts care of the lower one, whether the lower one is able to relate to the world, whether it can sense pain, or the like.

When examining these questions we might ask how robots fit into the scheme.

The book has a very short and defective bibliography presented as notes.

44. **Shear, J.,** editor, *Explaining Consciousness – 'The Hard Problem'* MIT Press, (1995).

This is a collection of philosophical papers written by professional philosophers and scientists attending a conference that sought to establish a scientific basis for the study of consciousness. The book is easily readable by the general reader and is structured as a debate between the various participants at the conference. All of the discussions are pre-scientific – they set out various hypotheses of what consciousness might be and make very tenuous claims about how consciousness might be studied scientifically. The book does, however, provide a basis for further argument.

45. **Shoemake, K.** "Arcball Rotation Control" in *Graphics Gems IV* pp. 175-192, Academic Press, (1994).

This article is intended for the strong programmer who is sufficiently familiar with algebra to learn the properties of quaternions. The article does, however, contain full source code, so mathematical ability is not a prerequisite.

The Arcball is a computer interface where the user uses a mouse to drag a point displayed on the surface of a sphere. The Arcball has the unusual, but extremely useful, property that mouse strokes are additive, despite the fact that rotations are not commutative. Two Arcballs can be combined in a simple way to describe 4D rotations.

The interface is reported to be highly intuitive, so it may well provide a mathematical model of the human perception of rotation.

46. **Spalter, A.M.** *The Computer in the Visual Arts* Addison-Wesley, (1999).

This book explains computer graphics devices and techniques in enough detail for the artist to understand the medium of computer graphics and how to use its techniques to produce pictures and, to a lesser extent, 3D sculptures and animation. A bibliography cites scientific papers and text books on computer graphics and art. There is also a list of web sites that show computer art, provide chat rooms, and give advice to the visual artist who uses computers as a medium.

47. **Spiegel, M.R.** *Schaum's Outline Series: Theory and Problems of Vector Analysis,* McGraw Hill, (1974).

This text book aims to teach vector algebra to students with a modest mathematical background. It has very many worked examples at various levels of difficulty. There is no bibliography.

48. **Stolfi, J.** *Orientable Projective Geometry* Academic Press, (1991).

This book is aimed at the strong programmer and seeks to popularise a form of projective geometry that involves a double covering of space. In one copy of space all objects are oriented clockwise and in the other copy anti-clockwise. The standard projective operations of *meet* and *join* are developed so as to preserve the orientability of objects across multidimensional sub-spaces. This allows computer graphics software to treat the surface of a high dimensional object consistently, and so draw it and shade its surface correctly. Stolfi recommends the use of Plücker co-ordinates as an alternative to homogeneous co-ordinates, especially in high dimensional spaces. He also provides a multidimensional generalisation of the cross-ratio.

There is a short bibliography citing scientific papers and text books in computer science and geometry.

49. **Swartz, R.J.,** editor, *Perceiving, Sensing, and Knowing* Anchor Books, (1965).

This compendium of philosophical essays is accessible by the general reader. It provides an introduction to the main themes of philosophising about the nature

of perceptions and what we can know of the world. All of the essays are in the analytical tradition of philosophy, that is, they analyse the way that people talk about their perceptions and knowledge. A bibliography groups philosophical text books and papers by subject.

50. **Turing, A.M.** *On Computable Numbers, With an Application to the Entscheidungsproblem* Proceedings of the London Mathematical Society, series 2, volume 42, part 3, pages 230 – 265, (1936).

This paper is aimed at the professional logician, but the early part, which defines the Turing machine, can be understood by the general reader who is prepared to work through the examples given and, possibly, to consult modern introductory texts on the Turing machine. The paper ends with difficult proofs using functional calculus. See next.

51. **Turing, A.M.** *On Computable Numbers, With an Application to the Entscheidungsproblem. A Correction.* Proceedings of the London Mathematical Society, series 2, volume 43, part 7, pages 544 – 546, (1937).

This is a correction to the proofs above.

52. **Turing, A.M.** *Computing Machinery and Intelligence* Mind, volume LIX, number 236, pages 433–460, (1950).

This highly readable paper is accessible to the general reader. It provides an easy introduction to the Turing machine and proposes the now famous Turing test to provide an empirical answer to the question, “Can computers think?” Modern reports of this test have been changed to meet contemporary concerns with political correctness. These reports miss much of the content of the paper, especially where it touches on the issue of the human body, sex, society, and religion.

I strongly recommend reading this paper.

53. **Watson, G.** editor, *Free Will* Oxford University Press (1989). First published 1982.

This is a collection of papers by professional philosophers on the problem of free will as it relates to determinism and morality.

54. Weihrauch, K. & Zhong, N. "Is Wave Propagation Computable or can Wave Computers Beat the Turing Machine?" *Proceedings of the London Mathematical Society,* number 85, part 3, pp. 312-332, (2002).

This mathematical paper re-examine Pour-El and Richards' analysis of the wave equation[39]. Weihrauch and Zhong confirm that the equation is incomputable using Pour-El and Richards' analysis; but they use a stronger definition of computability, and a stronger physical theory, to show that the wave equation is computable in all cases that are physically measurable. They conclude that there is almost no possibility of using the wave equation to create a super-Turing machine.

55. Yandell, K.E. *Philosophy of Religion* Routledge, (2002).

This text book is aimed at the general reader who has some knowledge of formal logic and is willing to consider the religions of the world and put them to the test of logical consistency. Some of the religions analysed fail this test.

There is a wide ranging and interesting chapter on free will and an interesting characterisation of religion. Religions are said to be diagnoses of spiritual ills along with a prescription for their cure. It is suggested that whilst humans are generally prone to the same kinds of spiritual ills, a human can have some, all, or none of these ills, so that an individual might be in the need of the ministrations of many religions. This is an argument for inter-faith dialogue or the philosophical study of religion.

This book is particularly pertinent to the role of religion for robots, especially if we create robots, by accident or design, that suffer our spiritual ills.

A bibliography cites philosophical papers and text books, as well as translated sources of religious texts.

Index

$\vec{x}$ $\vec{y}$

U

V

W

Z